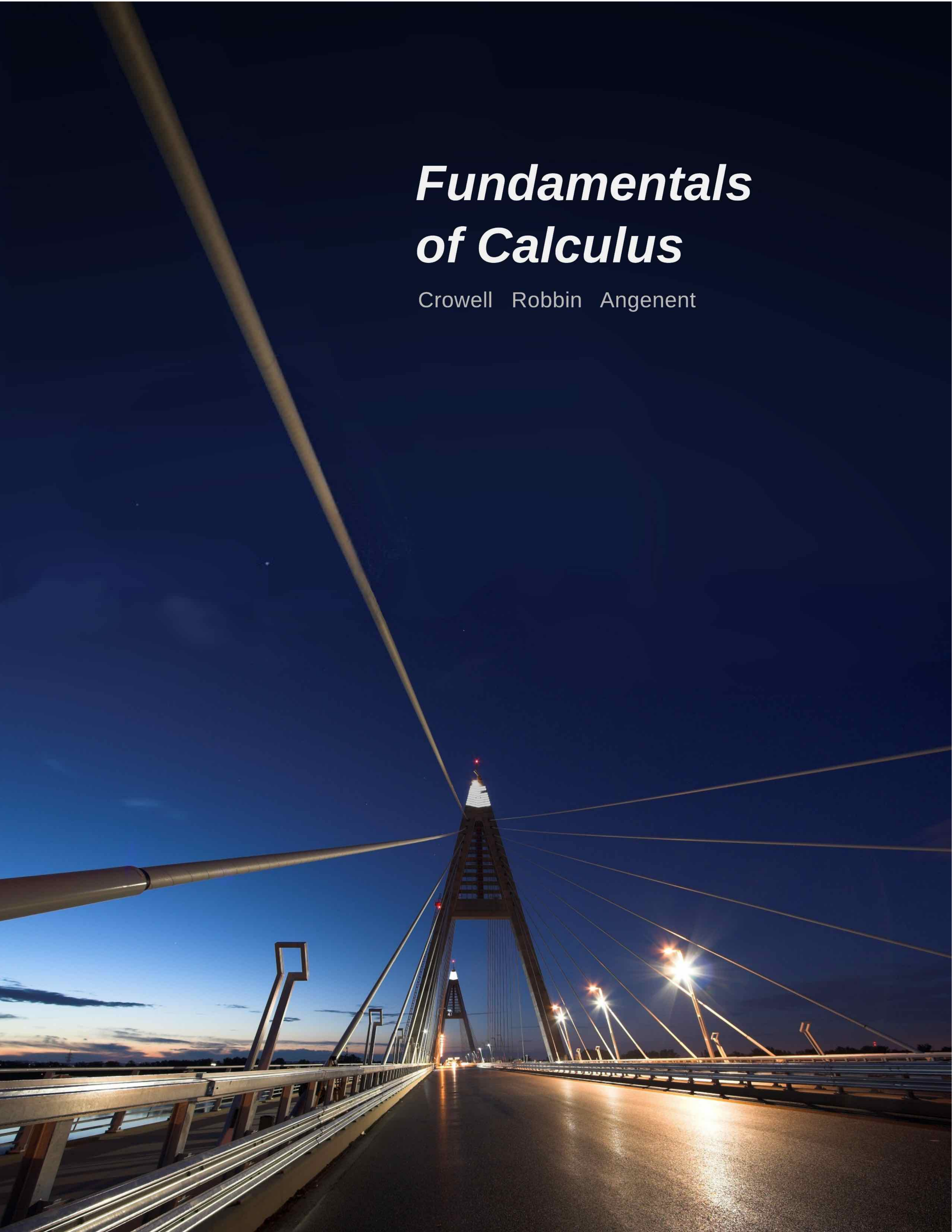


Fundamentals of Calculus

Crowell Robbin Angenent

Fundamentals of Calculus

This open-source book was produced by Benjamin Crowell and combines his work with material from a preexisting open-source book by Joel Robbin and Sigurd Angenent, www.math.wisc.edu/undergraduate/calculus-instructors-page.

www.lightandmatter.com

Light and Matter
Fullerton, California
www.lightandmatter.com

rev. January 18, 2016

Brief Contents

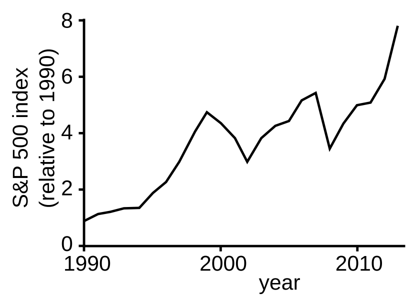
Contents

Chapter 1

An informal introduction to the derivative

1.1 Review: functions and the slope of a linear function

Calculus is the study of rates of change, and of how change accumulates. For example, figure a shows the changes in the United States stock market over a period of 24 years. The y axis of this graph is a certain weighted average of the prices of stock, and the x axis is time, measured in years. This is an example of the concept of a mathematical *function*, which you've learned about in a previous course. We say that the stock index is a function of time, meaning that it depends on time. What makes this graph the graph of a function is that a vertical line only intersects it in one place. This means that at any given time, there is only one value of the index, not more than one.



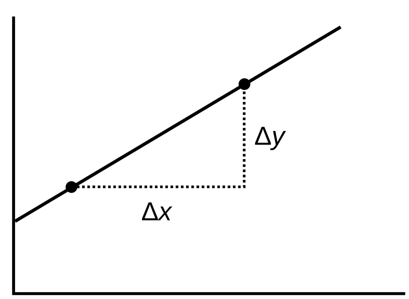
a / The S&P 500 stock index is a function of time.

Figure a shows a function that was determined by measurement and observation, but functions can also be defined by a formula. For example, we could define a function y by stating that for any number x, the value of the function is given by $y(x) = x^2$. We sometimes state this kind of thing more casually by referring to "the function $y = x^2$" or "the function x^2."

I drew figure a by graphing yearly data, so it's made of line segments that connect one year to the next. Each of these line segments has a *slope*, defined as

$$\text{slope} = \frac{y_2 - y_1}{x_2 - x_1}. \tag{1}$$

The slope measures how fast the function is changing. A positive slope says the function is increasing, negative decreasing. If the slope is zero, the function is not changing at all.



b / Given two points on a line, we can find its slope by computing $\Delta y/\Delta x$, the rise over the run.

It's often convenient to express this kind of thing with the notation Δ, the capital Greek letter delta, which is the equivalent of our Latin "D" and here stands for "difference." In terms of this notation, we have

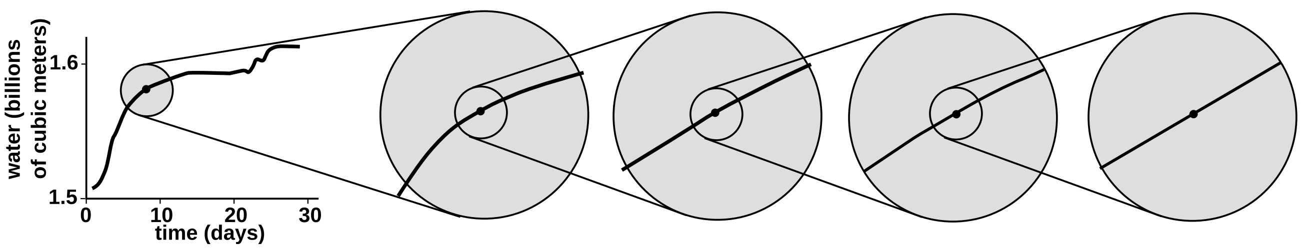
$$\text{slope} = \frac{\Delta y}{\Delta x}. \tag{2}$$

A symbol like Δy indicates the *change* in y, $\Delta y = y_2 - y_1$. It doesn't mean a number Δ multiplied by a number y.

1.2 The derivative

1.2.1 An informal definition of the derivative

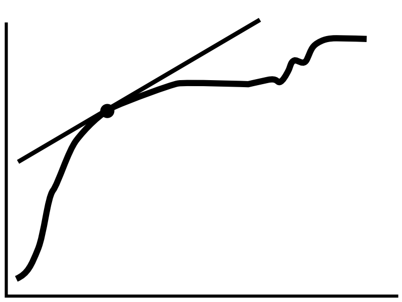
In many real-world applications, it makes sense to think of change as occurring smoothly and continuously. For example, the level of water in a reservoir rises and falls with time. Although it's true that this change happens one molecule at a time, so that in theory there are abrupt jumps, these jumps are too tiny to matter in practice.



c / The original graph, on the left, shows the water level in Trinity Lake, California, for the thirty-day period beginning March 7, 2014. Each successive magnification to the right is by a factor of four.

We want to keep track of the net rate of flow into the reservoir. We would like to define this rate as the slope of the graph, but the graph isn't a line, so how do we do that? We could pick two points on the graph and connect them with a line segment, but that would only represent an average rate of flow, not the actual rate of flow as it would be measured by a flow gauge at one particular time.

To get around these difficulties, we imagine picking a point of interest on the graph and then zooming in on it more and more, as if through a microscope capable of unlimited magnification. As we zoom in, the curviness of the graph becomes less and less apparent. (Similarly, we don't notice in everyday life that the earth is a sphere.) In figure c, we zoom in by 400%, and then again by 400%, and so on. After a series of these zooms, the graph appears indistinguishable from a line, and we can measure its slope just as we would for a line. This is an intuitive description of what we mean by the slope of a function that isn't a line. We call this slope the *derivative* of the function at the point of interest. This is admittedly not a mathematically rigorous definition, but it fixes our minds on the concept we want. A useful example is that if we consider a car's odometer reading as a function of time in hours, then its speedometer reading is the derivative of the odometer reading.



d / The tangent line at a point on a curved graph.

If we were only shown the ultra-magnified view in the rightmost part of figure c, we wouldn't know that the graph was curved at all. We would think the whole thing was a line. This hypothetical line is called the *tangent line* at the point marked with a dot. When you stand on the earth's surface and look at a point on the horizon, your line of sight is a tangent line to the surface. The derivative of a function is the slope of the tangent line.

1.2.2 Locality of the derivative

From this informal definition it seems that the derivative of a function at a certain point should depend only on the behavior of the function near that point, not far away. To state this idea precisely, we need to use some notation referring to sets, reviewed in box 1.1, and intervals.

Often it is useful to define a set of all the real numbers that lie within a certain range, between numbers a and b. This is called an interval. We can define intervals that contain or don't contain their endpoints.

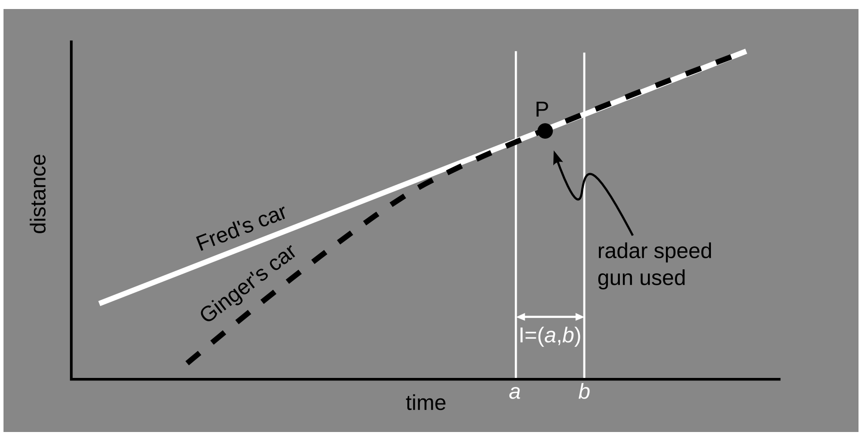
Definition

type of interval	*definition*	*abbreviation*
closed	$\{x \vert x \geq a \text{ and } x \leq b\}$	$[a, b]$
open	$\{x \vert x > a \text{ and } x < b\}$	(a, b)

We can also have intervals like $[a, b)$ and $(a, b]$, which are defined in the obvious way. A similar notation for infinite intervals is introduced in problem i4, p. 41.

Locality of the derivative

The derivative is *local*, in the following sense. Suppose there is an interval $\mathrm{I} = (a, b)$ on which the functions f and g are equal. That is, for any $x \in \mathrm{I}$, $f(x) = g(x)$. Then at any point in I, the derivatives of f and g are the same.



e / Fred and Ginger are both driving on the freeway. As Ginger is about to pass Fred, she notices a motorcycle cop, so she abruptly decelerates and then stays alongside Fred. The derivative of their position is their speed. The derivative is local, so by the time the cop measures their speeds, at point P, they are the same.

▷Box 1.1 **Sets**

A *set* is a collection of things. The things can, for example, be numbers. They can even be other sets. A set can be defined by listing the things it holds, which are called its elements or members. For example, the solutions of the equation $x^2 = 1$ are the members of the set $\{-1, 1\}$. Often we deal with infinite sets such as the set of all the natural numbers, and it is then impossible to list all the elements. Instead, we can define a set using notation like this:

$$\mathrm{S} = \{x | x^2 > 0\},$$

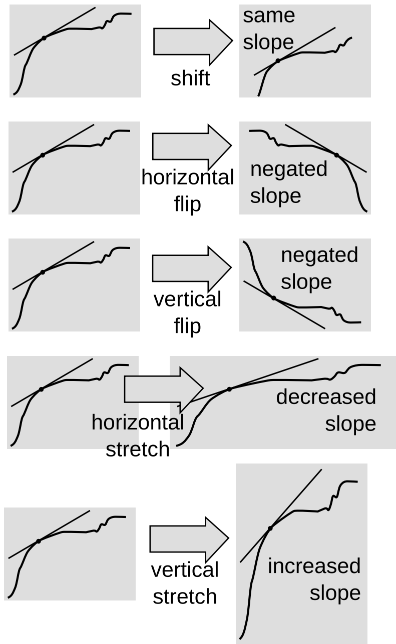
read as, "the set of all x such that x squared is greater than zero." Often, as in this example, we don't explicitly say what to consider as the *possible* values of x; since the focus of calculus is on real numbers, the implication in this course is usually that "the set of all x such that ..." means "the set of all *real numbers* x such that ..."

The notation $\in$ means "is a member of," e.g., $1 \in \mathrm{S}$ for the set S defined above.

Two sets are the same if they have the same members. For example, let

$$\mathrm{T} = \{a | a^2 > 0\} \quad \text{and}$$
$$\mathrm{U} = \{g | g \neq 0\}$$

Because S, T, and U have the same members, they are equal, $\mathrm{S} = \mathrm{T} = \mathrm{U}$.



f / Some properties of the derivative.

1.2.3 Properties of the derivative

The following properties of the derivative are intuitively reasonable based on our conceptual definition, and they will be enough to allow us to do quite a bit of interesting calculus before we come back and make a more general definition.

constant The derivative of a constant function is zero.

line The derivative of a linear function is its slope.

shift Shifting a function $y(x)$ horizontally or vertically to form a new function $y(x+a)$ or $y(x)+b$ gives a derivative at any newly shifted point that is the same as the derivative at the corresponding point on the unshifted graph.

flip Flipping the function $y(x)$ horizontally or vertically to form a new function $y(-x)$ or $-y(x)$ negates its derivative at corresponding points.

addition The derivative of the sum of two functions is the sum of their derivatives.

stretch Stretching a function $y(x)$ vertically to form a new function $ry(x)$ multiplies its derivative by r at the corresponding points, while stretching it horizontally to make $y(x/s)$ divides its derivative by s.

no-cut Suppose that for a certain point P on the graph of a function, there is a unique linear function ℓ that passes through P but doesn't cut through the graph at P. Then the graph of ℓ is the tangent line, and the derivative of the function at P equals the slope of the line.

As an example of the stretch rule, cars sold in the U.S. have odometers that read out in units of miles, while those sold elsewhere are calibrated in kilometers, so their readings are greater by the conversion factor $r = 1.6$. By the stretch property, cars outside the U.S. also have speedometer readings that are greater by this factor: they read out in *kilometers* per hour.

There is usually, but not always, a line like the one described by the no-cut property. Sometimes there is a tangent line but it isn't a no-cut line. If this kind of mathematical puzzle interests you, try sketching the graphs of the functions x^3 and $\sqrt{x}$. You should be able to convince yourself that their tangent lines at $x = 0$ can't be described by no-cut functions.

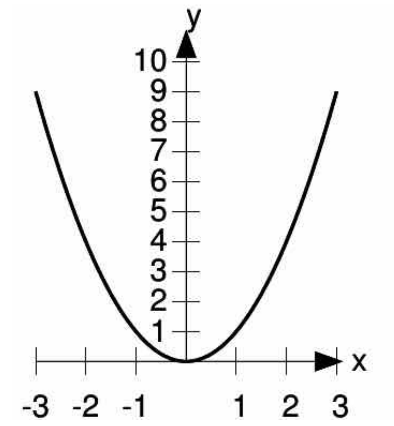
By the way, these are just names I've given to these properties, and if you use them with other people, they won't know what you mean. Once we've done more calculus, we'll see that several of these properties are actually special cases of a more general rule called the chain rule.

▷Box 1.2 **Ideas about proof: stating your assumptions**

The properties listed here can be used to solve problems, as in section 1.2.4, where we'll calculate the derivative of the function $y = x^2$. But math isn't just calculation. We also want to *prove* general facts. A proof always requires certain starting assumptions, e.g., you can't prove to a friend that cap-and-trade is the best way to deal with global warming if your friend won't admit that global warming exists. This list of properties includes enough assumptions to prove quite a few general facts about derivatives.

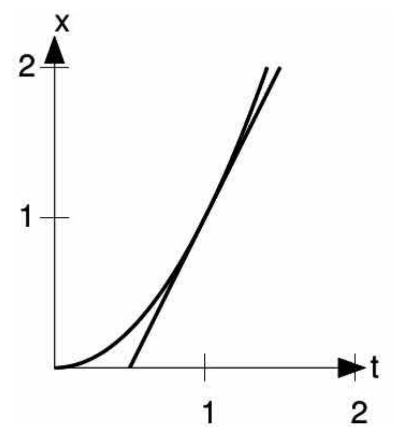
1.2.4 The derivative of the function $y = x^2$

As our first example of a derivative, let's use the function $y = x^2$. Its graph is a parabola. The simplest point at which to find its derivative is $x = 0$, the central point of the graph. From figure g, it seems like zooming in more and more on this point would give something that looked more and more like a horizontal line, and this suggests that the derivative at this point is zero. We can confirm this by using the flip property. Flipping the graph horizontally across the y axis doesn't change the graph. (Recall that a function with this symmetry is called an *even* function.) Since the flip doesn't change the function, it can't change the derivative of the function. But the flip rule says that when we flip a function, the derivative is negated at the corresponding point on the new graph. Here the point of interest is $x = 0$, and that point doesn't move when we flip it, so its corresponding point on the new graph is the same point. Thus the derivative at $x = 0$ must be the same as itself, but also equal to minus itself. Zero is the only number that remains the same when we reverse its sign, so the derivative at the center of the graph is zero.



g / The function $y = x^2$.

How about the derivative at the point $x = 1$? Here we can apply the no-cut rule. By laying a ruler against this point, we find that the linear function $\ell(x) = 2x-1$ seems to intersect the parabola without cutting across it. To prove that this is true, we can compute the difference between the two functions, $y(x) - \ell(x) = x^2 - 2x + 1$. Completing the square allows us to rewrite this as $(x - 1)^2$, which is clearly positive for any value of x other than 1. Therefore the function ℓ meets the conditions of the no-cut rule, and the derivative of x^2 at $x = 1$ is 2.



h / The line $2x - 1$ intersects the function x^2 without cutting it.

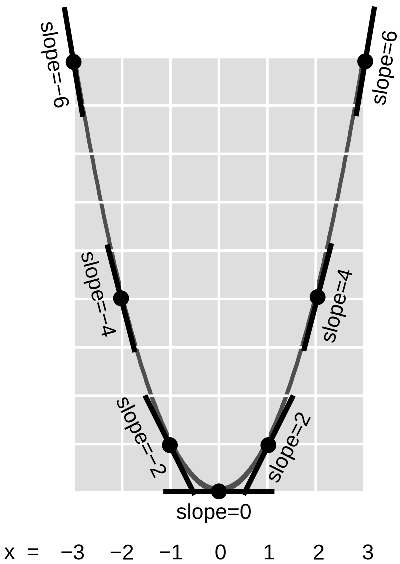
Having found the derivative of x^2 at $x = 1$, we can now use the stretch rule to find it at any other point. For example, suppose we want to know the derivative at $x = 3$. If we were to take the graph of the function x^2 and stretch it by a factor of 3 horizontally and 9 vertically, we would get the same graph again. These stretches take the point $(1, 1)$, where we know the derivative, to the point $(3, 9)$, where we want to know it. The stretch rule tells us that the horizontal stretch decreases the derivative to $1/3$ of its original value, but the vertical stretch increases it by 9 times, so that over all, the derivative at $(3, 9)$ is $(1/3)(9) = 3$ times greater than its value at $(1, 1)$. Thus the derivative at $x = 3$ equals 6.

There is nothing special about the number 3. The method that we applied to $x = 3$ would work for any other number x, not just for 3. We find that the derivative of the function x^2 at any point x equals $2x$. Taking stock of what we've done, we started with the function x^2, and found that at any point x, the derivative was $2x$.

1.2.5 The derivative of a function is a function itself.

We've found that the derivative of the function x^2 at a point x equals $2x$. The expression $2x$ can be thought of as a function of x. So what we've really done is to take a function and construct a new function that gives the derivative of the original function at each point. One way of notating this new function is y', read "y prime." We have

$$y = x^2$$
$$y' = 2x.$$

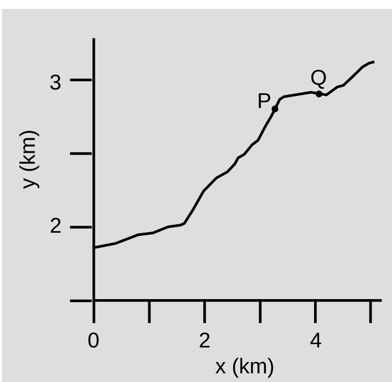


i / The derivative of x^2 is itself a function. As we change x, the slope of the tangent line changes.

The craft of finding this kind of derivative-function from the original function is called *differentiation.* We have differentiated the function x^2 and gotten its derivative, the function $2x$.

Hiking Example 1

Figure j shows a graph of my favorite route for climbing a mountain near where I live. (My wife rolls her eyes when I tell her the dog and I are doing this hike yet again.) How steep is the hike? There is no generic answer to this question, since the derivative of this function is itself a function. The derivative depends on x, so it has different values in different places. The slope of the graph at point P appears to be the steepest, with $y' \approx 0.80$. At other points, y' has smaller values. At Q, it's slightly negative. The derivative y' is a *function* of x; it depends on which part of the hike you're presently climbing.

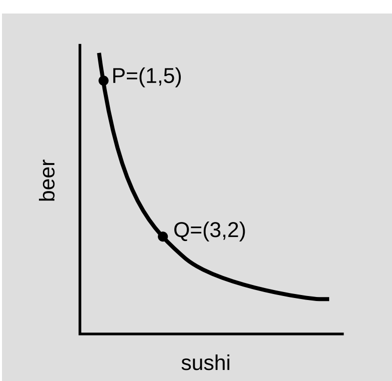


j / Example 1.

An indifference curve Example 2

Let's say you enjoy beer, and you also enjoy sushi. How much would you prefer to have of each? Economists define a graph, figure k, called an *indifference curve*. For a particular person, any two points on the curve are supposed to be equal in preference; the person is indifferent as to which one they get. For example, the person whose indifference curve is drawn in figure k is equally happy having one piece of sushi and five beers, or having three pieces of sushi and two beers.

There is a quantity called the marginal rate of substitution (MRS), which is defined as minus the slope of the indifference curve, $-y'$. At point P in figure k, the MRS is high, which means that the person would have been just as happy to have another piece of sushi and a *lot* less beer. The MRS, $-y'$, is a *function* of where you are on the curve. If the person is at point Q on the graph, they have a moderate amount of beer and a moderate amount of sushi, so they consider them of more comparable value. Indifference curves are discussed further in section 3.4.3, p. 89.



k / Example 2.

What if x is in the exponent rather than the base? Example 3

The method used above to differentiate x^2 was basically a trick, and it depended on a special property of the function x^2, which is that its graph can be stretched horizontally and vertically in such a way that it can be brought back on top of itself again. The reason that this subject is called "calculus" rather than "trickery" is that we will soon (in ch. 2) develop more systematic methods for *calculating* rates of change — methods that don't depend on tricks.

It may nevertheless be of interest to note that a similar trick is capable of telling us something about a different type of function, one in which x appears in the exponent rather than the base. What about the function 2^x, for example? A pair of rabbits marches off of Noah's ark. Two bunnies become four, then 8, 16, 32, and so on. What is the derivative of this function, i.e., the rate of change of the rabbit population per generation? (Strictly speaking, the derivative is only meaningful if we fill in all the non-integer values of x, which isn't really meaningful in terms of rabbits, since you can't have a fraction of a rabbit.)

It happens that the function 2^x, like x^2, can be brought back on top of itself again in a simple geometrical way. Instead of a horizontal stretch and a vertical stretch, we use a horizontal shift and a vertical stretch. For example, if we shift the graph of 2^x to the right by 3 units, and then stretch it vertically by a factor of 8, we get back the same graph again. This has come about because of the more fundamental property of exponential functions $b^{c+d} = b^c b^d$. (In our example, the base b is 2.) As a result, we find that after 3 generations, when the rabbit population goes up by a factor of 8, its derivative *also* goes up by a factor of 8. That is, the derivative of an exponential function $y = b^x$ is proportional to y, or

$$y' = (\ldots)y,$$

where "..." is a constant of proportionality that depends on the base b. What is the constant of proportionality? We'll return to this question in example 6 on p. 51.

A similar example is credit card debt. The more credit card debt you have, the faster your debt grows; in this example, the constant of proportionality relates to the interest rate.

Discussion question

A What is wrong with the logic of the following argument? *You should believe in God, because if you don't, when you die you'll go to Hell.*

Refer to box 1.2 on p. 16.

1.3 Derivatives of powers and polynomials

In section 1.2.4, we found that the derivative of x^2 was $2x$. Straightforward application of the same technique to x^3 gives $3x^2$. We see a pattern:

Derivatives of powers
The derivative of x^n equals nx^{n-1}, if n is any integer greater than or equal to 1.

Observing the pattern or giving examples is not enough to prove this general rule (box 1.3). To prove this for all these values of n, rather than carrying out the proof for one value at a time, it will be more convenient to use techniques developed later in the book (section 2.6, p. 57).

If we combine this with the addition and stretch rules, we know enough to differentiate any polynomial.

Differentiating a polynomial Example 4

▷ Find the derivative of $y = x^3 - 7x + 1$.

▷ The addition property of the derivative tells us that we can break this problem down into three parts,

$$(x^3 - 7x + 1)' = (x^3)' + (-7x)' + (1)',$$

where the primes indicate "derivative of ..." The stretch property says that $(-7x)'$ is the same as $(-7)(x)'$, so the derivative of our polynomial becomes

$$(x^3)' + (-7)(x)' + (1)'.$$

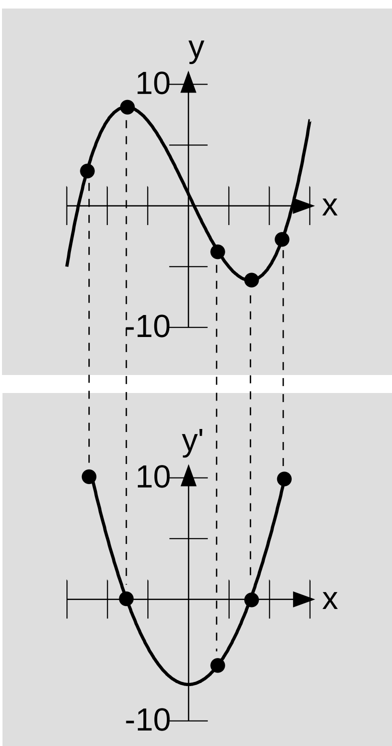
We know how to differentiate powers: $(x^3)' = 3x^2$, $(x') = 1$, and $(1)' = 0$. (We could have found the second term from the line property, and the final one from the constant property.) The result is

$$y' = 3x^2 - 7.$$

The functions y and y' are graphed in figure l, and five points are marked as examples of how the *slope* of y corresponds to the *value* of y'. Reading across from left to right on the top graph, the slopes are positive, zero, negative, zero, and positive. On the bottom graph, the values of y' are easily seen to be positive, zero, negative, zero, and positive.

▷Box 1.3 **Ideas about proof: examples don't prove a rule**

An example can't prove a general rule. French is the official language of Côte d'Ivoire, but that doesn't prove that it's the official language of all of Africa. In fact there are other countries in Africa, such as Egypt, that speak different languages, such as Arabic. In general, an example can never prove a general rule, but a *counterexample* (Egyptians speaking Arabic) can *disprove* a rule (all of Africa speaking French).



l / Example 4. The top graph shows the original function, the bottom its derivative.

1.4 Two trivial hangups

1.4.1 Changing letters of the alphabet

The following point is relatively trivial, but nevertheless hangs up many students in applying calculus to real life. In a calculus textbook, we typically use the letters x and y, with y being a function of x. That is, x is the independent variable, and y is the dependent one. In real-life applications, however, the variables have definite meanings, and we want to use letters that make it easy to remember what they stand for.

For example, suppose that a social media company has a certain number of users, and they need to have enough computing power at their data center to be able to handle all of those users. This computing power will cost them a certain amount of money per month. In this example, it would be natural to use the notation u for the number of users, and c for the monthly cost in dollars. Then c depends on u, and we have a function $c(u)$. Let's say the function is this:

$$c = u^2$$

This is not an unrealistic equation to imagine for this example, since the company has to keep track of every user's relationship to every other user. For example, user Andy may be able to mark himself as a "fan" or "follower" of user Betty, and then the company has to store a piece of information in a database to record this relationship. If there are a thousand users, there are 1000×1000 or a million such possible relationships that may need to be stored in a database.

Now if the company's user base is growing, it's of interest to them to know how much their costs will go up for each additional user (the marginal cost). This would be expressed by the derivative $c'(u)$. Although the letters of the alphabet are different than the ones we used in our earlier examples, that makes no difference in how we do the math. If differentiating $y = x^2$ with respect to x gives $y' = 2x$, then differentiating $c = u^2$ with respect to u gives the same result but with the letters changed,

$$c' = 2u$$

1.4.2 Symbolic constants

The vertical stretch property of the derivative tells us that if we know a derivative such as

$$(x^2)' = 2x,$$

then we can differentiate a function like $5x^2$ by simply letting the factor of 5 "come along for the ride,"

$$\begin{aligned}(5x^2)' &= (5)(x^2)' \\ &= (5)(2x) \\ &= 10x.\end{aligned}$$

Now suppose that we want to differentiate bx^2, where b is a constant, i.e., b doesn't depend on x. To many students this looks like a much more difficult and abstract problem, but the procedure is the same:

$$\begin{aligned}(bx^2)' &= (b)(x^2)' \\ &= (b)(2x) \\ &= 2bx.\end{aligned}$$

The same goes for a vertical shift. If we aren't intimidated by computing

$$(x^2+5)' = (x^2)' = 2x,$$

then there is no reason to be scared of the similar computation (again with b being a constant) of

$$(x^2+b)' = (x^2)' = 2x.$$

1.5 Applications

1.5.1 Velocity

Defining velocity

One of our prototypical examples has been the odometer and speedometer on a car's dashboard. In fact, if we want to *define* what velocity means, we have to define it as a derivative. Suppose an object (it could be a car, a galaxy, or a subatomic particle) is moving in a straight line. By choosing a unit of distance and a location that we define as zero, we can superimpose a number line onto this line. (In the example of the car, the unit of distance might be kilometers, and the zero position would be the point on the road at which we last pushed the button to zero the odometer.) Let the position defined in this way be x. Then x is a function of time t (such as the time measured on a clock), and we notate this function as $x(t)$. Note that although we typically use the letters x and y in a generic mathematical context, with y being a function of x, in our present example it is more natural to use different letters, and now x is the *dependent* variable, not the independent one. That is, x is a function of t, but t may not be a function of x; for example, if a car stops and backs up, then it can visit the same position twice, so that a graph of t versus x would fail the vertical line test for a function. In this notation, the velocity v is defined as the derivative

$$v(t) = x'(t).$$

Constant acceleration

An important special case is the one in which the position function is of the form

$$x(t) = \frac{1}{2}at^2,$$

where a is a constant, and the factor of $1/2$ is conventional, and convenient for reasons that will become more apparent in a moment. Differentiating with respect to t, we have the velocity function

$$v(t) = at,$$

where the symbolic constant a has been treated like any other constant, and the $1/2$ in front has been canceled by the factor of 2 that comes down from the exponent. We see that the velocity is proportional to the amount of time that has passed. If t is measured in seconds and v in meters per second (m/s), then the constant a, called the *acceleration*, tells us how much speed the object gains with every second that goes by, in units of m/s/s, which can be written as $\mathrm{m/s^2}$. Falling objects have an acceleration of about $9.8\ \mathrm{m/s^2}$. This is a measure of the strength of the earth's gravity near its own surface.

Dropping a rock down a well Example 5

▷ Looking down into a dark well, you can't see how deep it is. If you drop a rock in and hear it hit the bottom in 2 seconds, how deep is the well?

▷

$$x(t) = \frac{1}{2}at^2 \approx 20\ \mathrm{m}$$

The shift property applied to constant acceleration Example 6

The equations for constant acceleration were given above with the unstated assumption that both the position and the velocity would be zero at the time $t = 0$. If we relax this assumption, then the position function can be of the more general form

$$x(t) = x_o + \frac{1}{2}a(t - t_o)^2,$$

where t_o is some initial time, at which the position equals x_o. By the shift property of the derivative (p. 16), the velocity function is then

$$v(t) = a(t - t_o).$$

1.5.2 When do you need a derivative?

Finding velocity from position data is a classic application of calculus, and yet how do we know when we really need calculus for this application? After all, many people do simple computations involving velocity without knowing calculus.

Here's an example where calculus really is required. In July 1999, Popular Mechanics carried out tests to find which car sold by a major auto maker could cover a quarter mile (402 meters) in the shortest time, starting from rest. Because the distance is so short, this type of test is designed mainly to favor the car with the greatest acceleration, not the greatest maximum speed (which is irrelevant to the average person). The winner was the Dodge Viper, with a time of 12.08 s. If we divide the distance by the time, we get

$$v = \frac{\Delta x}{\Delta t} = 33.3 \text{ m/s},$$

which is about 74 miles per hour or 120 kilometers per hour. Not a very impressive speed, is it? That's because it's wrong. During those twelve seconds of acceleration, the car didn't have just one speed. It started at a velocity of zero and went up from there. The top speed was nearly double the one calculated above (53 m/s $\approx$ 119 mi/hr $\approx$ 191 km/hr). The important point here is that when we measure a rate of change using an expression of the form

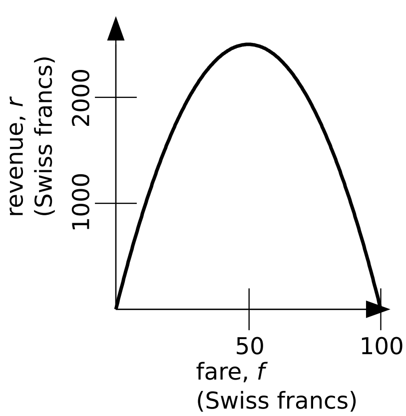
$$\frac{\Delta \ldots}{\Delta \ldots},$$

we only get the right answer if the rate of change is *constant*. In this example the rate of change is the velocity, and the velocity is not constant. To find the correct velocity, we first need to decide at which time we want to know the velocity, and then evaluate the derivative at that time.

1.5.3 Optimization

An extremely important use of the derivative is in optimization. For example, suppose that the operators of a privately owned mountain tram in Switzerland want to optimize their profit from transporting sightseers to a mountain summit in the Alps. The cost of building the tram is a sunk cost, and operating it for one day costs the same amount of money regardless of the number of passengers. Therefore the only goal is to get the maximum number of Swiss francs in the cash registers at the end of each day. The operators can raise the fare f in order to make more money, but if the fare is too high then not as many people will be willing to pay it. Suppose that the number of riders in a given day is given by $a - bf$, where a and b are constants. That is, if the ride was free, a passengers would ride each day, but for every one-franc increase in the fare, b people will decide not to go. The tram's daily revenue is then found by multiplying the number of riders by the fare, which gives the function

$$r(f) = (a - bf)f. \tag{3}$$



m / Revenue from a tram as a function of the fare charged.

For insight into what's going on, figure m shows this function in the case where $a = 100$ and $b = 1$. When the fare is zero, we get plenty of customers every day, but they don't pay anything, so our revenue is zero. When the fare is 100 francs, the number of paying passengers goes down to zero, so again we have no revenue.

Somewhere in between these extremes we have the fare that would optimize our revenue: the *maximum* of the function r. At this point on the graph, the derivative is zero, so to find it, we should differentiate r, set it equal to zero, and solve for f.

We haven't yet learned enough of the techniques of calculus to know how to find the derivative of a function with the form of equation (3), but by multiplying out the product we can make it into a polynomial, which is a form that we do know how to differentiate:

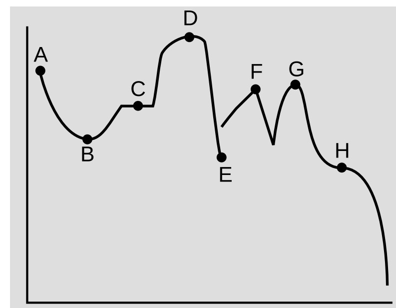
$$\begin{aligned} r(f) &= -bf^2 + af \\ r'(f) &= -2bf + a \end{aligned}$$

Setting r' equal to zero, we have

$$\begin{aligned} 0 &= -2bf + a \\ f &= \frac{a}{2b}. \end{aligned}$$

With the particular numerical values used to construct the graph, this gives an optimal fare of 50 francs, which looks about right from the graph.

By searching for points where the derivative is zero we can often, but not always, find the the points where a function takes on its maximum and minimum values. The term extremum (plural extrema) is used to refer to these points. Figure n shows that quite a few different things can happen, and that searching for a zero derivative doesn't always tell us the whole story. We have a zero derivative at point G, but G is only a maximum compared to nearby points; we call G a local maximum, as opposed to the global maximum D. The zero-derivative test doesn't distinguish a local minimum like B from a local maximum. A zero derivative may not indicate a local extremum at all, as at C and H. We can have points such as E and F where the derivative is undefined. An extremum can occur at a point like A that is the endpoint of the function's domain.[1] We will come back to these technical points in more detail later in the book.[2]



n / A zero derivative often, but not always, indicates a local extremum. Sometimes we have a zero derivative without a local extremum, and sometimes a local extremum with an undefined or nonzero derivative.

o / The railroad tracks stretch toward a vanishing point at infinity. Are there infinitely big or infinitely small numbers?

1.6 Review: elementary properties of the real numbers

I began this chapter by defining calculus as the study of rates of change, but it could equally well be described as the study of infinity. The intuition behind the derivative is that we zoom in on a selected point on a smooth curve, until the curve appears like a line and we can measure the slope of the line. But the curve won't appear perfectly straight until we've cranked up our microscope to an *infinitely big* magnification, at which point we'll be seeing values

[1]For a more thorough review of notions such as the domain of a function, see section 5.5, p. 131.

[2]section 3.4.1, p. 86

of Δx and Δy that are *infinitely small* (but not zero). Calculus was invented by Isaac Newton and Gottfried Wilhelm von Leibniz back in the era of powdered wigs and silk stockings, and in those days the concept of "number" was still in the process of being standardized and formalized.[3] Newton and Leibniz found it convenient to work with symbols representing infinitely big and infinitely small numbers, and a debate ensued about whether it was all right to call those things "numbers."

Today we think about this kind of thing in a different way. Decisions about what to allow as a legal number are thought of not as matters of right and wrong but as definitions. We define certain sets of numbers, including:

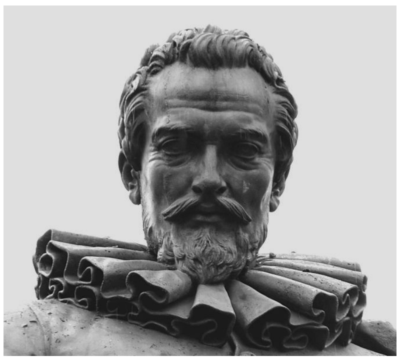
the integers: whole numbers such as -1, 0, and 1

the rational numbers: ratios of integers such as 2/1 and 3/4

the real numbers, including quantities like π and $\sqrt{2}$

the complex numbers, such as $\sqrt{-1}$

Do these systems contain infinitely big and infinitely small numbers? Can they? Should they?



1.41421...

1⓪4①1②4③2④1 ...

p / Simon Stevin (1548-1620) was a Flemish mathematician and engineer who lived a century before the invention of the calculus. He wrote a book on decimals, using a notation somewhat different from the modern one. (The figure shows the modern notation and Stevin's notation for the decimal expansion of $\sqrt{2}$.) Stevin's decimals represent an alternative approach to defining what we mean by a real number: rather than defining them by listing their properties, we can define them by *constructing* them out of simpler objects (decimal digits). Stevin argued for allowing any arbitrary, infinite string of digits, which is equivalent to including all the real numbers but forbidding infinitely big and infinitely small numbers.

To answer these questions, we need to give a more definite account of how these number systems are defined. One good way to define them is with a list of their axioms. (For an alternative, constructive approach, see figure p.) Here is a list of axioms for the system of real numbers. Except as otherwise stated, each of these properties holds for *any* real-number values of the symbols x, y, ...

commutativity $x + y = y + x$ and $xy = yx$

identities There exist numbers 0 and 1 such that for any x, $x+0 = x$ and $1x = x$.

inverses For any x, there exists a number $-x$ such that $x+(-x) = 0$. For any nonzero x, there exists $1/x$ such that $(x)(1/x) = 1$.

associativity $x + (y + z) = (x + y) + z$ and $x(yz) = (xy)z$

distributivity $x(y + z) = xy + xz$

ordering We can define whether or not $x < y$, and this ordering relates to the addition and multiplication operations in specific ways, which you've seen defined in a previous course on algebra and which for brevity we will not explicitly give here.

[3] For more on the history, see Blaszczyk, Katz, and Sherry, "Ten misconceptions from the history of analysis and their debunking," `arxiv.org/abs/1202.4153`.

This list of axioms holds for the real numbers, but it fails for the integers, since for example the integer 2 doesn't have an inverse that is an integer. It also fails for the complex numbers, which don't have a well-defined ordering. The list seems detailed and precise, so it may come as a surprise that it does *not* suffice to prove anything about whether or not infinite numbers exist. The list of axioms is in fact not enough to characterize the real numbers. Later in this book we will add another axiom, called the completeness axiom (section 4.5, p. 111), to the list. The completeness axiom holds for the reals but not the rationals, and it also rules out the existence of infinitely large or infinitely small real numbers. It is possible to extend the real number system to a larger one that does include infinities (section 2.9, p. 64).

1.7 The Leibniz notation

1.7.1 Motivation

Lacking the more precise modern ideas described in section 1.6, Leibniz argued as follows. Let's just make Δx and Δy infinitely small (but not zero). In modern terminology, this means that they can't be real numbers. To make it clear that we're talking about infinitely small differences in x and y, we change the notation to $\mathrm{d}x$ and $\mathrm{d}y$. Recall that Δ is the Greek version of capital "D," so we're using a smaller version of the letter, "d," to represent a change that is smaller (in fact, infinitely small). Dividing these two "numbers" (whatever mysterious species of number they may turn out to be), we get the derivative,

$$\frac{\mathrm{d}y}{\mathrm{d}x}.$$

Although the notation's original justification was not up to modern standards of rigor, it is one of the most expressive and well-designed mathematical notations ever devised, and has been the most commonly used notation for the derivative ever since Leibniz published it in 1686. Around 1970, mathematicians clarified some of these issues and essentially justified and codified the centuries-old procedures for manipulating the $\mathrm{d}y$'s and $\mathrm{d}x$'s; section 2.9 on p. 64 boils these modern developments down to a simple set of practical rules.

q / Gottfried Wilhelm Leibniz (1646-1716).

1.7.2 With respect to what?

One of the good things about the Leibniz notation is that it states clearly what we're differentiating *with respect to.* For example, $\mathrm{d}v/\,\mathrm{d}t$ could indicate how much a car was speeding up with each passing second of time, while $\mathrm{d}v/\,\mathrm{d}x$ would measure the speed gained with each meter that it moved down the road.

1.7.3 Shows units

Another selling point of the notation is that it shows the units of the derivative. For example, the definition of velocity, expressed in Leibniz notation, is

$$v = \frac{\mathrm{d}x}{\mathrm{d}t}.$$

On the left-hand side we have velocity, whose units in the SI are meters per second. On the right we have a tiny change in position, which has units of meters, divided by a tiny change in time, which has units of seconds. In terms of units, then, the equation reads as

$$\mathrm{m/s} = \frac{\mathrm{m}}{\mathrm{s}},$$

which works out correctly. In more complicated examples, checking the units like this is a powerful method for checking your answer to a calculus problem.

▷Box 1.4 **The SI**

The metric system is the system of units used universally in engineering and the sciences, as well as in daily life in every country except the United States. Formally known as the Système International (SI), it was invented during the French Revolution. For mechanical (as opposed to electrical) measurements, the SI uses three basic units:

meters for length

kilograms for mass

seconds for time

Other measurements are built from these, e.g., meters per second (m/s) for velocity.

There is a system of prefixes that represent powers of ten in which the exponent is a multiple of three. The most common of these are kilo- $= 10^3$, and milli- $= 10^{-3}$. (The prefix centi- $= 10^{-2}$ is used only in the centimeter, and doesn't require memorization since we know that dollars and euros are subdivided into 100 cents.)

Burning gasoline Example 7

▷ Let x be a car's odometer reading and g the amount of gasoline burned since the odometer was zeroed. One can think of x as a function of g. Many cars have a digital display that shows the function $x'(g)$ in real time. Express this using the Leibniz notation. What is the interpretation of this derivative, and what units does it have?

▷ The Leibniz notation is $\mathrm{d}x/\mathrm{d}g$, which makes it clear that the units are kilometers per liter, km/L (or, in U.S. units, miles per gallon). The interpretation is that this number gives a measure of how efficient the car is at using fuel to transport you a given distance.

An insect pest Example 8

▷ An insect pest from the United States is inadvertently released in a village in rural China. The pests spread outward at a rate of s kilometers per year, forming a widening circle of contagion. Find the number of square kilometers per year that become newly infested. Check that the units of the result make sense. Interpret the result.

▷ Let t be the time, in years, since the pest was introduced. The radius of the circle is $r = st$, and its area is $a = \pi r^2 = \pi(st)^2$. To make this look like a polynomial, we have to rewrite it as $a = (\pi s^2)t^2$. The derivative is

$$\begin{aligned}\frac{\mathrm{d}a}{\mathrm{d}t} &= (\pi s^2)(2t)\\ &= (2\pi s^2)t\end{aligned}$$

The units of s are km/year, so squaring it gives $\mathrm{km}^2/\mathrm{year}^2$. The 2 and the π are unitless, and multiplying by t gives units of $\mathrm{km}^2/\mathrm{year}$, which is what we expect for $\mathrm{d}a/\,\mathrm{d}t$, since it represents the number of square kilometers per year that become infested.

Interpreting the result, we notice a couple of things. First, the rate of infestation isn't constant; it's proportional to t, so people might not pay so much attention at first, but later on the effort required to combat the problem will grow more and more quickly. Second, we notice that the result is proportional to s^2. This suggests that anything that could be done to reduce s would be very helpful. For instance, a measure that cut s in half would reduce $\mathrm{d}a/\,\mathrm{d}t$ by a factor of four.

A whirling bucket — Example 9

▷ Figure r shows a bucket full of water that is being whirled rapidly, so that the water spreads out from the center. The surface of the water forms a parabola with the equation

$$y = \frac{x^2}{c},$$

where c is a constant. Infer the units of c, find the slope of the water's surface, and check the units of your answer.

▷ Both x and y are measured in units of meters, so we have

$$\mathrm{m} = \frac{\mathrm{m}^2}{\text{units of } c}.$$

If the units of the left and right sides are to be equal, c must have units of meters as well.

Differentiation gives the slope of the water's surface as

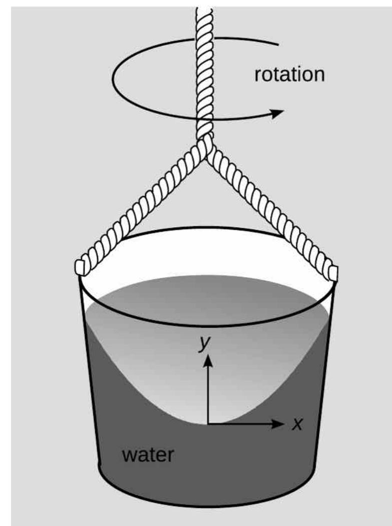
$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{2x}{c},$$

where the factor of $1/c$ "comes along for the ride," as with any multiplicative constant.

Checking the units of the result, we have

$$\frac{\mathrm{m}}{\mathrm{m}} = \frac{(\text{unitless}) \cdot \mathrm{m}}{\mathrm{m}},$$

which checks out.



r / Example 9.

1.7.4 Operator interpretation

Sometimes the Leibniz notation gives an unwieldy, top-heavy tower of symbols:

$$\frac{\mathrm{d}\left(\frac{x^2}{2}+\frac{1}{7}\right)}{\mathrm{d}x} = x$$

One way to avoid this awkwardness is to revert to the "prime" notation:

$$\left(\frac{x^2}{2}+\frac{1}{7}\right)' = x$$

But a more common solution is write the function being differentiated over on the right:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^2}{2}+\frac{1}{7}\right)=x$$

This can be seen simply as a typographical expedient, or it can be given a mathematical interpretation: we can think of $\frac{\mathrm{d}}{\mathrm{d}x}$ as meaning "take the derivative of," in the same way that $\sqrt{}$ means "take the square root of." We call $\frac{\mathrm{d}}{\mathrm{d}x}$ the *operator* describing the operation of taking a function and giving back the function that is its derivative. Math teachers who dislike the historical connotations of the Leibniz notation in terms of infinitely small numbers will sometimes present the operator interpretation as the *only* correct interpretation, but such a prescription robs the student of some of the utility of the notation, e.g., by making it impossible to do the kind of reasoning shown in example 8.

1.8 Approximations

We saw in section 1.5.2 on p. 23 that the derivative can't be calculated as $\Delta y/\Delta x$ unless the derivative is constant, i.e., unless the function's graph is a line. In the Leibniz notation, this is

$$\frac{\mathrm{d}y}{\mathrm{d}x}\neq\frac{\Delta y}{\Delta x}.$$

But if we take two points very close together on a graph, then the curvature doesn't matter too much, and the line through those points is a good approximation to the tangent line, as in figure s. When then have the approximation

$$\frac{\mathrm{d}y}{\mathrm{d}x}\approx\frac{\Delta y}{\Delta x}.$$

It may be of interest to use either side of this as an approximation to the other.

1.8.1 Approximating the derivative

Suppose you can't remember that the derivative of x^2 is $2x$, but you need to find the value of the derivative at $x=1$. As in figure s, let point P be

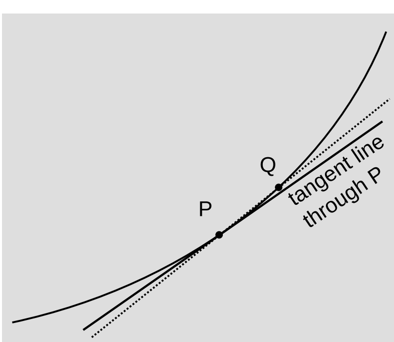
$$(1.0000, 1.0000),$$

and let Q be the nearby point

$$(1.0100, 1.0201).$$

We then have:

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}x}&\approx\frac{\Delta y}{\Delta x}\\&=\frac{1.0201-1.0000}{1.0100-1.0000}\\&=\frac{0.0201}{0.0100}\\&=2.01\end{aligned}$$



s / The dotted line through P and Q is a good approximation to the tangent line through P.

This is quite a good approximation to the exact answer, 2. If we needed a better approximation, we could take Q even closer to P. In reality we would use this technique in cases where we didn't know the exact answer, and we would then want to know how accurate our result was. To do this, we could redo the calculation with a smaller value of Δx, say 0.001, and look for the most significant decimal place that changed.

1.8.2 Approximating finite changes

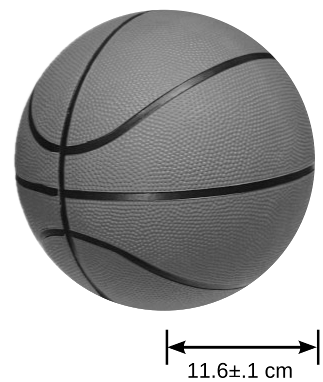
Sometimes we know the derivative and want to use it as an approximation to find out about finite changes in the variables. For example, the Women's National Basketball Association says that balls used in its games should have a radius of 11.6 cm, with an allowable range of error of plus or minus 0.1 cm (one millimeter). How accurately can we determine the ball's volume?

The equation for the volume of a sphere gives $V = (4/3)\pi r^3 = 6538\ \text{cm}^3$ (about six and a half liters). We have a function $V(r)$, and we want to know how much of an effect will be produced on the function's output V if its input r is changed by a certain small amount. Since the amount by which r can be changed is small compared to r, it's reasonable to apply the approximation

$$\frac{\Delta V}{\Delta r} \approx \frac{\mathrm{d}V}{\mathrm{d}r},$$

which gives

$$\begin{aligned} \Delta V &\approx \frac{\mathrm{d}V}{\mathrm{d}r}\Delta r \\ &= 4\pi r^2 \Delta r. \end{aligned}$$



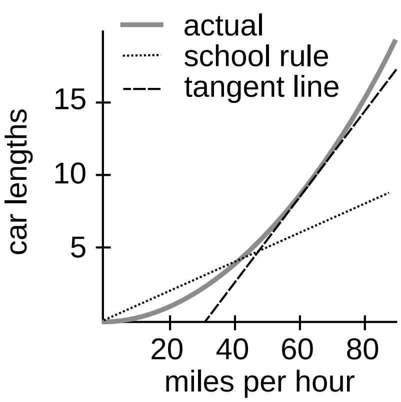
t / How accurately can we determine the ball's volume?

(Note that the factor of $4\pi r^2$ can be interpreted as the ball's surface area.) Plugging in numbers, we find that the volume could be off by as much as $(4\pi r^2)(0.1\ \text{cm}) = 170\ \text{cm}^3$. The volume of the ball can therefore be expressed as $6500 \pm 170\ \text{cm}^3$, where the original figure of 6538 has been rounded off to the nearest hundred in order to avoid creating the impression that the 3 and the 8 actually mean anything — they clearly don't, since the possible error is out in the hundreds' place.

This calculation is an example of a very common situation that occurs in the sciences, and even in everyday life, in which we base a calculation on a number that has some range of uncertainty in it, causing a corresponding range of uncertainty in the final result. This is called propagation of errors. The idea is that the derivative expresses how sensitive the function's output is to its input.

The example of the basketball could also have been handled without calculus, simply by recalculating the volume using a radius that was raised from 11.6 to 11.7 cm, and finding the difference between the two volumes. Understanding it in terms of calculus, however, gives us a different way of getting at the same ideas, and often allows us to understand more deeply what's going on. For example, we noticed in passing that the derivative of the volume was simply the surface area of the ball, which provides a nice geometric visualization. We can imagine inflating the ball so that its radius is increased by a millimeter. The amount of added volume equals the surface area of the ball multiplied by one millimeter, just as the amount of volume added to the world's oceans by global warming equals the oceans' surface area multiplied by the added depth.

As another example of an insight that we would have missed if we hadn't applied calculus, consider how much error is incurred in the measurement of the width of a book if the ruler is placed on the book at a slightly incorrect angle, so that it doesn't form an angle of exactly 90 degrees with spine. The measurement has its minimum (and correct) value if the ruler is placed at exactly 90 degrees. Since the function has a minimum at this angle, its derivative is zero. That means that we expect essentially no error in the measurement if the ruler's angle is just a tiny bit off. This gives us the insight that it's not worth fiddling excessively over the angle in this measurement. Other sources of error will be more important. For example, is the book a uniform rectangle? Are we using the worn end of the ruler as its zero, rather than letting the ruler hang over both sides of the book and subtracting the two measurements?



u / Stopping distance in car lengths, as a function of initial speed in miles per hour. The stopping distances were measured using professional drivers on a track. I've defined a car length as 4.8 meters, which is the length of a Honda Accord. The dotted line shows the traditional rule taught in schools in the US, one car length per 10 m.p.h. of speed. The dashed line is the tangent at 60 miles per hour, which is the best linear approximation for speeds near this one.

1.8.3 Linear approximation to a curve

Many people who, like me, learned to drive in the United States were taught that when following another car, we should leave space equal to one car length for every 10 miles per hour of speed. This rule has the advantage of being easy to compute in your head while you're on the freeway, but figure u shows that it's a poor approximation. This is an example of a situation that occurs over and over again in real life, which is that we would like to approximate a complicated nonlinear function using a simple linear one. The derivative is the slope of the tangent line, and the tangent line is the *best* possible line to approximate a given function near a particular point.

Here is a general procedure for finding the best linear approximation to a nonlinear function:

1. Pick some point on the graph that is near the center of the region for which we're interested in getting a linear approximation.

2. Differentiate the function to find the slope of the tangent line through this point.

3. Given a point on a line and the line's slope, we can find the equation of the line. One way to do this is to write down the definition of the slope as $\Delta y/\Delta x$.

Ice cream Example 10

▷ Fred drives an ice cream truck in Deadhorse, Alaska, where the average temperature in the summer is about 10 degrees Celsius. During the long Arctic winter nights, Fred has developed a mathematical model showing that his daily revenue y in dollars is related to the Celsius temperature x by the equation

$$y = -800 + 100x - x^2.$$

Find a useful linear approximation to this equation.

▷ Since the average temperature in summer is about $x = 10$, let's find the best linear approximation near this point. Differentiation gives $y' = 100 - 2x$, and plugging in $x = 10$ gives a slope

$$\frac{\Delta y}{\Delta x} = 80. \qquad \text{[slope of the tangent line]}$$

If we plug in the value $x = 10$ to the equation for y itself, we find that the point

$$(10, 100) \qquad \text{[a point on the tangent line]}$$

is the one that we're trying to find the tangent line through. We therefore have

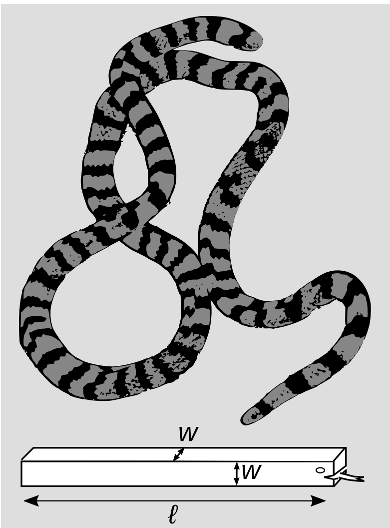
$$\frac{y - 100}{x - 10} = 80 \qquad \text{[point-slope form of the line]}$$

for the equation of the best linear approximation. Fred is interested in calculating his profits y, so he solves this for y to find $y = -700 + 80x$. As an approximation to the true (nonlinear) function, this is

$$y \approx -700 + 80x. \qquad \text{[slope-intercept form]}$$

1.9 More about units

In section 1.7.3 on p. 28, we briefly discussed the idea of checking your calculus by analyzing the units of measurement. If you had a good high school chemistry or physics course, you may have already learned how to do this to check your algebra. If not, then you may find it helpful to study this section, which lays out the ideas in more detail.



v / A snake approximated as a box.

Figure v shows a cute snake, along with its even cuter geometrical idealization as a rectangular box. The snake has

length ℓ, in units of meters (m)
width w, in units of meters (m),
mass M, in units of kilograms (kg).

(Some people would say "in units of length," and "in units of mass," but to be more concrete I'm using the SI units listed in box 1.4 on p. 28.)

It makes sense to manipulate these quantities in certain ways:

$$
\begin{array}{rl}
4w, & \text{the snake's waistline,} \\
w^2\ell, & \text{its volume in cubic meters (m}^3\text{),} \\
\dfrac{M}{w^2\ell}, & \text{its density in kg/m}^3\text{,}
\end{array}
$$

or

$$2w + \ell \le 1.14 \text{ m},$$

which tells us whether this snake is legal as carry-on luggage.

But some combinations don't make sense:

$$
\begin{array}{rl}
\ell + M & \text{can't add meters to kilograms} \\
w\ell = w^2\ell & \text{can't equate area to volume} \\
\cos M & \text{can't take the cosine of a mass}
\end{array}
$$

Some quantities are unitless. I have two dogs, and the 2 is a unitless 2; in general, a count is unitless. When we form a ratio between two numbers that have the same units, the result is unitless. For example, the rectangular snake in the figure has $\ell/w = 12.6$, which is unitless; one way to tell that it's unitless is that if we enlarge or reduce the drawing, the quantities that have units grow or shrink, but the proportions such as ℓ/w stay the same.

The following rules apply:

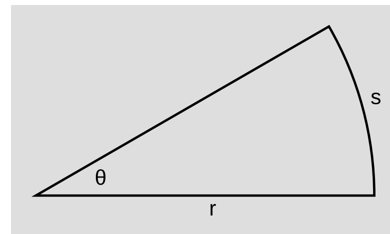
1. In addition, subtraction, and comparisons, all terms must have the same units.

2. When you multiply or divide numbers, multiply or divide their units as well.

3. All the functions on your calculator that go beyond grade-school arithmetic require a unitless input and give a unitless output. These functions include logs, exponentials, and trig functions, and are referred to collectively as transcendental functions (sec. 5.1.2, p. 126).

Radians aren't units Example 11

Using the notation shown in figure b, the radian measure of the angle θ is defined as s/r. The arc length s and radius r both have units of meters, so by rule 2 their ratio is unitless. Therefore radians are not really a unit. This is required by rule 3 so that we can use them as inputs to trig functions.



w / Example 11.

Cosine is unitless Example 12

The cosine is adjacent/hypotenuse, so it's unitless, as required by rule 3.

Frequency Example 13

The period T of a vibration is defined as time it takes to go through one cycle. The frequency is defined as $f = 1/T$, and by rule 2 it has units of 1/seconds or s^{-1} (also known as Hz).

Area, or volume? Example 14

▷ You remember that $4\pi r^2$ is the formula either for the volume of a sphere or for its surface area, but you can't remember which it is. Which one does it have to be based on units?

▷ The 4π is unitless. By rule 2, the expression $4\pi r^2$ thus has units of m^2, i.e., square meters, or area.

Square roots Example 15

A square root is not a transcendental function, so rule 3 doesn't apply to it. For example, our snake has a cross-sectional area $A = w^2$. We then have $w = \sqrt{A}$, and it's OK to feed the square root function a unitful input: $\mathrm{m} = \sqrt{\mathrm{m}^2}$.

No units in the exponent Example 16

▷ We can compute w^2, where w has units. Does that mean we can also calculate 2^w?

▷ No, because then $2^w = e^{\ln(2^w)} = e^{w \ln 2}$, but then the input to the exponential would have units, violating rule 3. I.e., the base-2 exponential is transcendental, just like the base-e flavor.

Radioactivity Example 17

▷ As a radioactive substance decays, the fraction of it that remains after time t is given by $f = e^{-t/k}$, where k is a constant. Infer the units of k.

▷ By rule 3, t/k must be unitless, so k is in seconds.

Review problems

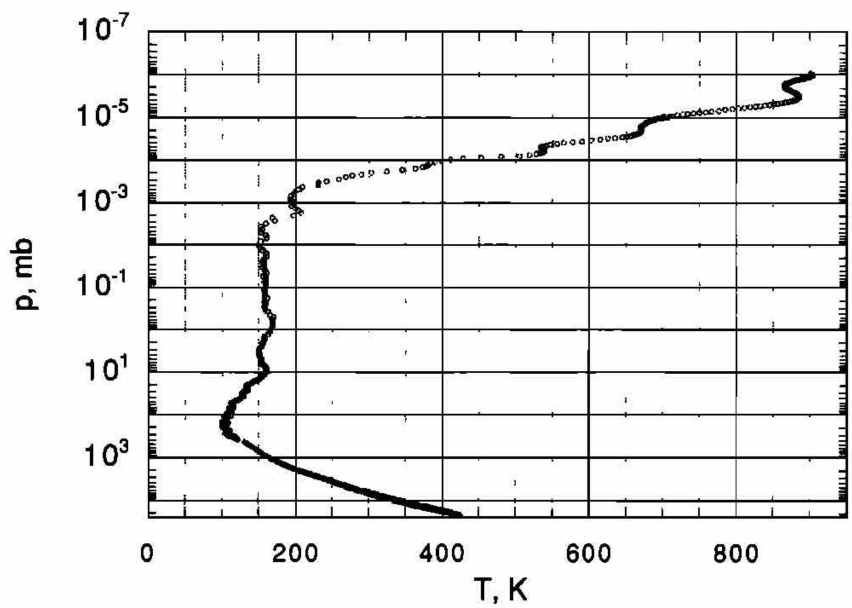
a1 A line with slope 3 passes through the point $(-7, 1)$. Find an equation for the line, solving for y. ✓

a2 A line passes through the points $(2, 3)$ and $(6, 5)$. (a) Find the slope. (b) Write an equation for the line, solving for y. ✓

a3 A line has the equation $4x - 3y + 1 = 0$. Find its slope. ✓

a4 A line has the equation $ax + by + c = 0$. If x changes by an amount Δx, find the amount Δy by which y changes. ✓

a5 The figure shows data on the pressure p and temperature T of the planet Jupiter, as measured by the Galileo probe in 1995. Can p be described as a function of T? Can T be described as a function of p? ▷ Solution, p. 224



Pressure (in millibars, mb) versus temperature (in degrees Kelvin, K) of the atmosphere of Jupiter, problem a5. For comparison, the atmospheric pressure and temperature at the earth's surface are about 1000 mb and 300 K. Although Jupiter is in the outer solar system and is in general very cold, the temperatures in its tenuous upper atmosphere are, counterintuitively, very hot; this feature of the graph is what would be referred to on earth as an "inversion layer." Seiff *et al.*, J. Geophys. Research 103 (1998) 22,857.

a6 Suppose that a line is expressed as an equation in the form $(\ldots)x + (\ldots)y + (\ldots) = 0$, where the $(\ldots)$ stand for constants. Under what conditions does y fail to be a function of x?

▷ Solution, p. 224

a7 Let x and y be real numbers. Which of these equations make y a function of x?

$$y = x \qquad y = x^2 \qquad x = y^2 \qquad y = x^3 \qquad x = y^3$$

▷ Solution, p. 224

a8 Let $\mathrm{S} = \{u | u^2 - 2u < 0\}$. Figure out what set of points is really being described here, and rewrite this as a simpler definition of the form $\mathrm{S} = \{\ldots | \ldots\}$. ▷ Solution, p. 224

Problems

c1 Differentiate the following functions with respect to t: $1,\ 7,\ t,\ 7t,\ t^2,\ 7t^2,\ t^3,\ 7t^3$. ▷ Solution, p. 224

c2 The functions f and g are defined by

$$f(x) = x^2 \text{ and } g(s) = s^2.$$

Are f and g the same function, or are they different?
▷ Solution, p. 225

c3 Let m be an amount of money. There are many examples from business, personal finance, and government in which it makes sense to imagine that m is a function of time, $m(t)$. Make up an example in which $m(t) = 0$ but $m'(t) \neq 0$. (Don't make up an equation, just explain a situation where this would happen and how it would be interpreted.) ▷ Solution, p. 225

c4 A seller offers something at a unit price P, and the quantity of units sold is Q. Ordinarily, we expect that P and Q would be related in some way that could be expressed by a graph, but there's no obvious way to decide which variable, P or Q, should be on which axis. The cause-and-effect relationship isn't clearly one way or the other: a change in price could cause a change in demand, but a change in demand could also prompt the seller to change the price. The graph is called the *demand curve.*

For some unusual goods, the demand is insensitive to the price. For example, the drug Soliris treats a genetic disease so rare that only about 8,000 people in the U.S. have it. The price P is about \$400,000 per patient per year. Since the benefits of treatment for these people are so great, and the cost is paid for by government or private insurers, changing P would not change Q. (a) How would this example look on a graph if we put P on the y axis and Q on the x axis? What if we did it the other way around? (b) In each case, discuss whether the graph is a function. (c) In each case, what can you say about the derivative based on the the informal definition given in section 1.2.1?

In problems d1-d5, a function is defined by giving an equation for y in terms of x. Find the derivative of the function.

d1 $y = 3x^4 - 2x^2 + x + 1$ ✓ ▷ Solution, p. 225

d2 $y = -7x^3 + x^2 - 7x - 7$ ✓

d3 $y = 2x^5 + 3x^4 - x^3 + 137$ ✓

d4 $y = 11x^{11} - 4x^4 + 2x - 8$ ✓

d5 $y = 3x^2 + 2x - 1$ ✓

Problems e2-e5 are each intended to be assigned randomly to one fourth of the students in a class.

e1 Differentiate $3z^7 - 4z^2 + 6$ with respect to z. Check your answer by picking an arbitrary value of z and applying the technique described in section 1.8.1, p. 30. ▷ Solution, p. 225

e2 Differentiate $4q^2 + 4q - 1$ with respect to q. Check your answer by the same technique as in problem e1. ✓

e3 Differentiate $-11w^3 + 5w^2 + 6$ with respect to w. Check your answer by the same technique as in problem e1. ✓

e4 Differentiate $c^{67} - 18c^2 + 987$ with respect to c. Check your answer by the same technique as in problem e1. ✓

e5 Differentiate $10r^{10} - 6r^6 + 7$ with respect to r. Check your answer by the same technique as in problem e1. ✓

e6 Find three different functions whose derivatives are the constant 7, and give a geometrical interpretation.
▷ Solution, p. 225

f1 Let the function y be defined by $y(x) = px^2 - qx + r$, where p, q, and r are constants. Find $y'(x)$. ✓

f2 Let the function h be defined by $h(u) = au^3 - \frac{u}{b} + c$, where a, b, and c are constants. Find $h'(u)$. ✓

In problems f3-f5 you will need to start by rewriting the given expressions in a form that you know how to differentiate. (If you've had some previous exposure to calculus, you may already know the product rule or chain rule. Some of these problems can be done using those rules, but they can also be done without them. If you use them, explain that you're doing so.)

f3 Let the function $f(x)$ be defined by $f(x) = (x+1)(2x+3)$. Find $f'(x)$. ✓

f4 Let the function q be defined by $q(c) = (2c^3)(7c)$. Find $q'(c)$. ✓

f5 Let the function z be defined by $z(j) = (aj)^4 - 7\left(\frac{j}{r}\right)^2$, where a and r are constants. Find $z'(j)$. ✓

f6 Let the function $f(x)$ be defined by

$$f(x) = \frac{x^{m+1}}{m+1},$$

where $m \neq -1$ is a constant. Find $f'(x)$. ✓

g1 Consider the function f defined by $f(x) = |x|$.
(a) Sketch its graph. If you're not sure what it would look like, try to gain insight by calculating points for a few values of x, including values that are positive, negative, and zero.
(b) On p. 14 I gave an informal definition of the tangent line and the derivative in terms of zooming in on a graph. Does this function have a well-defined tangent line at $x = 0$? A well-defined derivative?
(c) On p. 16 I defined a special type of tangent line called a no-cut line, and the definition requires that the no-cut line be unique, i.e., there is not more than one line with the given properties. Is there a no-cut line at $x = 0$ for this function?

g2 Consider the function f defined as follows:

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^2 & \text{if } x \geq 0 \end{cases}$$

(a) Sketch its graph. If you're not sure what it would look like, try to gain insight by calculating points for a few values of x, including values that are positive, negative, and zero.
(b) On p. 14 I gave an informal definition of the tangent line and the derivative in terms of zooming in on a graph. Does this function have a well-defined tangent line at $x = 0$? A well-defined derivative?
(c) On p. 16 I defined a special type of tangent line called a no-cut line, and the definition requires that the no-cut line be unique, i.e., there is not more than one line with the given properties. Is there a no-cut line at $x = 0$ for this function?

g3 Consider the function f defined by $f(x) = \sqrt{|x|}$.
(a) Sketch its graph. If you're not sure what it would look like, try to gain insight by calculating points for a few values of x, including values that are positive, negative, and zero. For insight, try a very small value of x such as 10^{-8}; think about how $f(x)$ compares with x for this small x, and what this tells you about the shape of the graph near $x = 0$.
(b) On p. 14 I gave an informal definition of the tangent line and the derivative in terms of zooming in on a graph. Does this function have a well-defined tangent line at $x = 0$? A well-defined derivative?
(c) On p. 16 I defined a special type of tangent line called a no-cut line, and the definition requires that the no-cut line be unique, i.e., there is not more than one line with the given properties. Is there a no-cut line at $x = 0$ for this function?

i1 Differentiate $at^2 + bt + c$ with respect to t.
[Thompson, 1919] ▷ Solution, p. 226

i2 Let the function f be defined by $f(x) = \frac{2}{3}x^2 + \frac{1}{5}x - \frac{5}{4}$. Find the value of x for which $f'(x) = \frac{2}{3}$. ✓

i3 The variables u and r are related by $u = \frac{5}{6}r^2 - \frac{1}{7}r + \frac{8}{3}$. Find the value of r that minimizes u. ✓

i4 Recall that the *range* of a function is the set of possible values its output can have. Find the ranges of the following functions.

$$\begin{aligned} f(x) &= 2x^2+3 \\ g(x) &= -2x^2+4x \\ h(x) &= 4x+x^2 \\ k(x) &= 1/(1+x^2) \\ \ell(x) &= 1/(3+2x+x^2) \\ m(x) &= 4\sin x+\sin^2 x \end{aligned}$$

(For m, if you've forgotten your trig you may wish to review from section 5.3, p. 128. It is possible to do this problem without knowing how to differentiate the sine function.)

You will find it convenient to express some of your answers using notations such as $[17,\infty)$, which is a standard way of extending the normal notation for finite intervals (p. 15) to describe infinite ones. This example means, as you'd imagine, the set $\{u|u \ge 17\}$. Although ∞ isn't a real number, the notation gets the idea across. The use of the) rather than a] is to show that there isn't a member of the set whose value is infinite.

Although you may be able to guess some of the answers by constructing a graph, that does not constitute a proof of the exact result.

i5 Consider the following four functions:

$$\begin{aligned} f(x) &= x^2-2x+\pi \\ g(u) &= u^{18}-2u^9+\pi \\ h(v) &= \ln(v^2-2v+\pi) \\ k(w) &= \tan^2 w-2\tan w+\pi \end{aligned}$$

Determine the minimum value of each function.

Although you may be able to get approximations to the answers by graphing, that does not constitute a proof of the exact result, which is what is required here. You may, however, find it helpful to *check* your exact results using graphing, e.g., on the online graphing app at desmos.com.

If you've forgotten some of your precalculus mathematics, you may wish to review trig from section 5.3, p. 128 and logarithms from section 5.7, p. 134. It is possible to do this problem without knowing how to differentiate the functions ln and tan; instead, reason about how the inputs and output of the functions work, and think about how the construction of functions h and k relates them to functions f and g. $\checkmark$

k1 Children grow *up*, but adults more often grow in the horizontal direction. Suppose we model a human body as a cylinder of height h and circumference c. The person's body mass is given by $m = \rho v$, where v is the volume and ρ (Greek letter rho, the equivalent of Latin "r") is the density. Find $\mathrm{d}m/\,\mathrm{d}c$, the rate at which body mass grows with waistline, assuing constant height and density. Check that your answer has the right units, as in example 8 on p. 28 and section 1.9 on p. 34. ✓

k2 Let t be the time that has elapsed since the Big Bang. In that time, one would imagine that light, traveling at speed c, has been able to travel a maximum distance ct. (In fact the distance is several times more than this, because according to Einstein's theory of general relativity, space itself has been expanding while the ray of light was in transit.) The portion of the universe that we can observe would then be a sphere of radius ct, with volume $v = (4/3)\pi r^3 = (4/3)\pi(ct)^3$. Compute the rate $\mathrm{d}v/\,\mathrm{d}t$ at which the volume of the observable universe is increasing, and check that your answer has the right units, as in example 8 on page 28 and section 1.9 on p. 34. Hint: We're differentiating with respect to t, and the thing being cubed is not just t, so this is not a form that you know how to differentiate. Use algebra to convert it into a form that you do know how to handle. ✓

k3 Kinetic energy is a measure of an object's quantity of motion; when you buy gasoline, the energy you're paying for will be converted into the car's kinetic energy (actually only some of it, since the engine isn't perfectly efficient). The kinetic energy of an object with mass m and velocity v is given by $K = (1/2)mv^2$.
(a) As described in box 1.4 on p. 28, infer the SI units of kinetic energy.
(b) For a car accelerating at a steady rate, with $v = at$, find the rate $\mathrm{d}K/\,\mathrm{d}t$ at which the engine is required to put out kinetic energy. $\mathrm{d}K/\,\mathrm{d}t$, with units of energy over time, is known as the *power*. Hint: We're differentiating with respect to t, and the thing being squared is not just t, so this is not a form that you know how to differentiate. Use algebra to convert it into a form that you do know how to handle. ✓
(c) Check that your answer has the right units, as in example 8 on page 28 and section 1.9 on p. 34.

m1 Section 1.2.3 on p. 16 defines the addition and vertical stretch properties of the derivative. If we assume that the addition property is true, prove that the vertical stretch property must hold for any stretch factor r that is a natural number (1, 2, 3, ...).
▷ Solution, p. 226

m2 Section 1.2.3 on p. 16 defines the constant and line properties of the derivative. Prove that the constant property follows from the line property.

m3 Section 1.2.3 on p. 16 defines the addition, constant, and vertical shift properties of the derivative. If we assume that the addition and constant properties are true, prove that vertical shift property must hold.

m4 An even function is one with the property $f(-x) = f(x)$. For example, $\cos x$ is an even function, and x^n is an even function if n is even. An odd function has $f(-x) = -f(x)$. Use the horizontal flip property of the derivative (p. 16) to prove that the derivative of an even function is odd.

n1 Rancher Rick has a length of cyclone fence L with which to enclose a rectangular pasture. Show that he can enclose the greatest possible area by forming a square with sides of length $L/4$. ▷ Solution, p. 226

n2 Prove that the total number of maxima and minima possessed by a third-order polynomial is at most two. ▷ Solution, p. 226

n3 A factory produces widgets, and the cost of production for a given year is $an + bn^2$, where n is the number produced, a is the basic cost of producing one widget, and b represents the fact that in order to increase volume, the factory must take expensive steps such as adding a night shift, paying overtime, or offering higher wages in order to attract more and better workers. The widgets are sold at a fixed unit wholesale price k, and there is unlimited demand.
(a) Find the optimal number of widgets that the factory should produce. √
(b) Check that your answer has the right units, as in example 8 on page 28 and section 1.9 on p. 34.
(c) Interpret the case where $b = 0$.
(d) Interpret the case where $k < a$.

n4 A steel sphere of radius r is dropped into an upright cylinder of radius $b \geq r$. For a fixed value of b, find the value of r that maximizes the amount of water that needs to be poured into the cylinder in order to cover the sphere. √

Problems p1-p3 are each intended to be assigned randomly to one third of the students in a class.

p1 A circle has area a, diameter d, and radius r. Express a in terms of r, d in terms of r, and a in terms of d. Find the derivatives $\mathrm{d}a/\mathrm{d}r$, $\mathrm{d}d/\mathrm{d}r$, and $\mathrm{d}a/\mathrm{d}d$. The Leibniz notation suggests that we should have

$$\frac{\mathrm{d}a}{\mathrm{d}r} = \frac{\mathrm{d}a}{\mathrm{d}d}\frac{\mathrm{d}d}{\mathrm{d}r}.$$

Is this actually true?

p2 A sphere has volume v, diameter d, and radius r. Express v in terms of r, d in terms of r, and v in terms of d. Find the derivatives $\mathrm{d}v/\mathrm{d}r$, $\mathrm{d}d/\mathrm{d}r$, and $\mathrm{d}v/\mathrm{d}d$. The Leibniz notation suggests that we should have

$$\frac{\mathrm{d}v}{\mathrm{d}r} = \frac{\mathrm{d}v}{\mathrm{d}d}\frac{\mathrm{d}d}{\mathrm{d}r}.$$

Is this actually true?

p3 An equilateral triangle has sides of length s, perimeter p, and area a. Express a in terms of p, p in terms of s, and a in terms of s. Find the derivatives $\mathrm{d}a/\mathrm{d}p$, $\mathrm{d}p/\mathrm{d}s$, and $\mathrm{d}a/\mathrm{d}s$. The Leibniz notation suggests that we should have

$$\frac{\mathrm{d}a}{\mathrm{d}s} = \frac{\mathrm{d}a}{\mathrm{d}p}\frac{\mathrm{d}p}{\mathrm{d}s}.$$

Is this actually true?

q1 As a tree grows in height h, it gains mass m, so that we have some function $m(h)$. If h is measured in units of meters, and m in kilograms, what are the units of the changes Δm and Δh and of the derivative $\mathrm{d}m/\mathrm{d}h$?

q2 A tank is filling with water. The volume (in cubic meters) of water in the tank at time t (seconds) is $V(t)$. What units does the derivative $V'(t)$ have?

r1 Use the technique in section 1.8.1 to obtain a numerical approximation to the derivative of the function $y = 1/(1-x)$ at $x = 0$. Find an answer accurate to three decimal places.

▷ Solution, p. 226

r2 Use the technique in section 1.8.1 to obtain a numerical approximation to the derivative of the function $y = \cos(x^3)$ at $x = 1$. Find an answer accurate to three decimal places. ✓

r3 Use the technique in section 1.8.1 to obtain a numerical approximation to the derivative of the function $y = \sin\sqrt{x}$ at $x = 1$. Find an answer accurate to three decimal places. ✓

r4 Use the technique in section 1.8.1 to obtain a numerical approximation to the derivative of the function $y = e^{\cos x}$ at $x = 1$. Find an answer accurate to three decimal places. ✓

r5 A function of the form $U = 1/(1+e^r)$ occurs in nuclear physics, and its derivative is interpreted as the force acting on a neutron or proton when it is at a distance r from the center of the nucleus. Use the technique in section 1.8.1 to obtain a numerical approximation to the derivative of this function at $r = 1$. Find an answer accurate to three decimal places. ✓

s1 Suppose that we measure a quantity x and compute from it $y = kx^n$, where k is a constant and n is a natural number. Let Δx be an estimate of the amount of possible measurement error in x, and let Δy be the corresponding error estimate for the output of the calculation.
(a) Show that if Δx is small compared to x, then

$$\frac{\Delta y}{y} \approx n \frac{\Delta x}{x}.$$

(b) Vernier calipers are used to measure the length of the sides of a square tile to a precision of 0.1%. Use the result of part a to find the possible error in an area computed from this length. ▷ Solution, p. 227

s2 A hobbyist is going to measure the height to which her model rocket rises at the peak of its trajectory. She plans to take a digital photo from far away and then do trigonometry to determine the height, given the baseline from the launchpad to the camera and the angular height of the rocket as determined from analysis of the photo. Comment on the error incurred by the inability to snap the photo at exactly the right moment. ▷ Solution, p. 227

s3 Joe sells square sheets of gold foil. Since gold is expensive, the sheets are sold by area a. If the area is too small, the customer gets upset, but if the area is too high, Joe is losing money. Therefore he wants to make sure that the area doesn't differ from a by more than Δa. In his shop, Joe marks off squares of length x.
(a) No measurement is perfectly exact. By what amount Δx can his length measurement be off if the resulting error in the area is to be no more than Δa? Use the approximation method described in section 1.8.2 on p. 31. √
(b) Check that your answer has the right units, as in example 8 on page 28 and section 1.9 on p. 34.
(c) If the desired area is $a = 4.000\ \mathrm{m}^2$, and the maximum allowable error in area is $0.001\ \mathrm{m}^2$, what is the biggest error Joe can afford to make when he marks off the length x? Express your result using an appropriate unit or in scientific notation, not as an awkward decimal with a string of zeroes. √

t1 (a) Let $y = x^p$, where the constant and p is a natural number. Find the best linear approximation to this function for values of x near 1. √
(b) Use the result of part a to approximate the value of 1.000001^{137} without a calculator. √

t2 The role of examples and counterexamples in proofs was introduced in box 1.3, p. 20. Sally claims that any function $y = x^n$, where n is a natural number, has $y' = 0$ at $x = 0$. To prove this, she gives a correct calculation of the derivative of $y = x^4$ at $x = 0$. (a) Explain why her proof is incorrect. (b) Disprove her claim by giving a counterexample.

t3 The role of examples and counterexamples in proofs was introduced in box 1.3, p. 20. The addition rule for the derivative (p. 16) tells us that the derivative of a sum is the sum of the derivatives. Huy proposes that the same thing holds for multiplication: that the derivative of a product is the same as the product of the derivatives. Disprove Huy's proposal by giving a counterexample.

Chapter 2

Limits; techniques of differentiation

In chapter 1 we started computing derivatives simply by appealing to a list of geometrically plausible properties (section 1.2.3, p. 16). These properties are true, and by taking them as axioms we were able to prove rigorously that, for example, the derivative of x^2 is $2x$ (section 1.2.4, p. 17). But there are many problems that are messy to solve by this limited toolbox of techniques, and many others for which we need qualitatively different tools.

Historically, the way Newton and Leibniz approached the problem was as follows. Suppose we want to take the derivative of x^2 at the point P where $x = 1$. We already know that we can get a good numerical approximation to this derivative by taking a second point Q, close to P, and evaluating the slope of the line through P and Q. (See section 1.8.1, p. 30). Now instead of picking specific numbers, let's just take point Q to lie at $x = 1 + \mathrm{d}x$, where $\mathrm{d}x$ is very small. Then the slope of the line through P and Q is

$$\begin{aligned}\text{slope of line PQ} &= \frac{\Delta y}{\Delta x} \\ &= \frac{(1+\mathrm{d}x)^2 - 1}{(1+\mathrm{d}x) - 1} \\ &= \frac{2\,\mathrm{d}x + \mathrm{d}x^2}{\mathrm{d}x}\end{aligned}$$

Now comes the crucial leap of faith, which mathematicians of later centuries began to feel was a little too sketchy. The number $\mathrm{d}x$ is supposed to be small, and when you square a small number you get an even smaller number. Since $\mathrm{d}x$ is supposed to be infinitely small, $\mathrm{d}x^2$ should be so small that it's utterly unimportant, even compared to $\mathrm{d}x$. Therefore we throw away the $\mathrm{d}x^2$ term and find that the slope of the tangent line is 2.

2.1 The definition of the limit

Starting in the 19th century, mathematicians became less and less satisfied with the logical justification for this style of doing calculus. The real number system had gradually become defined in a standardized way. It became clear that although one could have a number system that obeyed the axioms given in section 1.6 (p. 25)

▷Box 2.1 **Ideas about proof: proof by contradiction**

The practice of throwing away the square of $\mathrm{d}x$ shows that many mathematicians, for over a century, were willing to believe in nonzero numbers whose squares were zero. That contradicts what you learned in grade school, but it's not necessarily wrong. A proof has to be based on certain assumptions (box 1.2, p. 16). Those mathematicians simply didn't assume the same list of properties that is now standard for the real number system (section 1.6, p. 25).

Let's use those assumptions to prove that we can't have a nonzero x such that $x^2 = 0$. Suppose that such an x *did* exist. Then since $x \neq 0$, by the multiplicative inverse property there is a number $1/x$. Taking both sides of $x^2 = 0$ and multiplying by $1/x$ gives $x^2/x = 0/x$, or $x = 0$. But this contradicts the original claim that x was nonzero.

This is a *proof by contradiction*. If we assume something is true, and can then, through valid reasoning, arrive at mutually contradictory results, then the initial assumption must have been false.

and that included infinitely small numbers,[1] such a system would not be the same as the real numbers. Furthermore one would have a problem with the procedure of treating a dx^2 as if it were zero; one can prove from those axioms that zero itself is the only number whose square is zero (box 2.1, p. 47). For these reasons, mathematicians turned to a different way of defining the derivative, by using the new notion of a *limit*.

2.1.1 An informal definition

While it is easy to define precisely in a few words what a square root is ($\sqrt{a}$ is the positive number whose square is a) the definition of the limit of a function runs over several terse lines, and most people don't find it very enlightening when they first see it. So we postpone this momentarily and start by building up our intuition.

Definition of limit (first attempt)
If f is some function then

$$\lim_{x\to a} f(x) = L$$

is read "the limit of $f(x)$ as x approaches a is L." It means that if you choose values of x which are close *but not equal* to a, then $f(x)$ will be close to the value L; moreover, $f(x)$ gets closer and closer to L as x gets closer and closer to a.

The following alternative notation is sometimes used

$$f(x) \to L \quad \text{as} \quad x \to a;$$

(read "$f(x)$ approaches L as x approaches a" or "$f(x)$ goes to L is x goes to a".)

Example 1

If $f(x) = x + 3$ then

$$\lim_{x\to 4} f(x) = 7,$$

is true, because if you substitute numbers x close to 4 in $f(x) = x + 3$ the result will be close to 7.

Substituting numbers to guess a limit Example 2

What (if anything) is

$$\lim_{x\to 2} \frac{x^2 - 2x}{x^2 - 4}?$$

Here $f(x) = (x^2 - 2x)/(x^2 - 4)$ and $a = 2$.

We first try to substitute $x = 2$, but this leads to

$$f(2) = \frac{2^2 - 2\cdot 2}{2^2 - 4} = \frac{0}{0}$$

which does not exist. Next we try to substitute values of x close but not equal to 2. The table suggests that $f(x)$ approaches 0.5.

x	$f(x)$
3.000000	0.600000
2.500000	0.555556
2.100000	0.512195
2.010000	0.501247
2.001000	0.500125

Example 2.

[1] For more on this topic, see section 2.9 on p. 64.

Substituting numbers can suggest the wrong answer. Example 3

Our first definition of "limit" was not very precise, because it said "x close to a," but how close is close enough? Suppose we had taken the function

$$g(x) = \frac{101\,000x}{100\,000x + 1}$$

and we had asked for the limit $\lim_{x\to 0} g(x)$. Then substitution of some "small values of x," as shown in the table, could lead us to believe that the limit was 1. Only when you substitute even smaller values do you find that the limit is zero!

x	$g(x)$
1.000000	1.009990
0.500000	1.009980
0.100000	1.009899
0.010000	1.008991
0.001000	1.000000

Example 3.

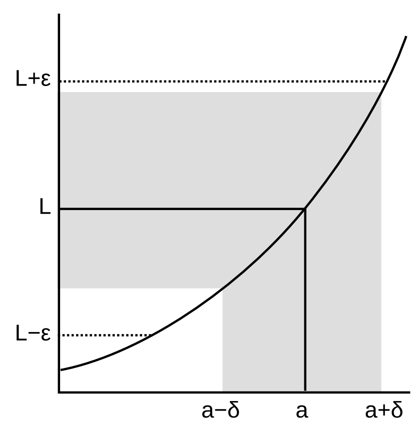
2.1.2 The formal, authoritative definition of the limit

The informal description of the limit uses phrases like "closer and closer" and "really very small." In the end we don't really know what they mean, although they are suggestive. Fortunately there is a better definition, i.e. one which is unambiguous and can be used to settle any dispute about the question of whether or not $\lim_{x\to a} f(x)$ equals some number L.

Definition of the limit

We say that L is the limit of $f(x)$ as $x \to a$, if the following two conditions hold:

1. The function $f(x)$ need not be defined at $x = a$, but it must be defined for all other x in some interval which contains a.
2. For every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all values of x in the domain of f with $|x - a| < \delta$, we have $|f(x) - L| < \varepsilon$.



a / The value of ε is imposed on us. We have succeeded in finding a value of δ small enough so that the outputs of the function do lie within the desired range. If we can do this for *every* value of ε, then the limit is L.

(The Greek letter "δ" is lowercase delta, equivalent to the Latin "d," and "ϵ" is epsilon, which is like Latin "e.")

Why the absolute values? The quantity $|x - y|$ is the distance between the points x and y on the number line, and one can measure how close x is to y by calculating $|x - y|$. The inequality $|x - y| < \delta$ says that "the distance between x and y is less than δ," or that "x and y are closer than δ."

What are ε and δ? The quantity ε is how close you would like $f(x)$ to be to its limit L; the quantity δ is how close you have to choose x to a to achieve this. To prove that $\lim_{x\to a} f(x) = L$ you must assume that someone has given you an unknown $\varepsilon > 0$, and then find a positive δ for which x values that close to a result in values of f that lie with the range the person has demanded. The δ you find will depend on ε.

Example 4

▷ Show that $\lim_{x\to 5} 2x+1 = 11$.

▷ We have $f(x) = 2x+1$, $a = 5$ and $L = 11$, and the question we must answer is "how close should x be to 5 if want to be sure that $f(x) = 2x+1$ differs less than ε from $L = 11$?"

To figure this out we try to get an idea of how big $|f(x) - L|$ is:

$$|f(x) - L| = \left|(2x+1) - 11\right| = |2x - 10| = 2 \cdot |x - 5| = 2 \cdot |x - a|.$$

So, if $2|x - a| < \varepsilon$ then we have $|f(x) - L| < \varepsilon$, i.e.

$$\text{if } |x - a| < \tfrac{1}{2}\varepsilon \text{ then } |f(x) - L| < \varepsilon.$$

We can therefore choose $\delta = \frac{1}{2}\varepsilon$. No matter what $\varepsilon > 0$ we are given our δ will also be positive, and if $|x - 5| < \delta$ then we can guarantee $|(2x+1) - 11| < \varepsilon$. That shows that $\lim_{x\to 5} 2x+1 = 11$.

Discussion question

A Figure a on p. 49 shows an example where δ is small enough for the given value of ε. What would the figure look like in a case where the value of δ was *not* small enough?

B Proof by contradiction was introduced in box 2.1 on p. 47. It can be considered as a specific mathematical version of an ancient technique of argument called *reductio ad absurdum*, or reduction to asburdity, which means to disprove something by showing that if it were true, then one could arrive at ridiculous results. When we say, "if that's true, then the Pope's not Catholic," we're implying that we could give a *reductio ad absurdum*. Suppose that Johnny insists on the obvious axiomatic truths (1) that monsters live under beds and inside closets; and (2) that monsters come out of their hiding places when the lights are turned out. Johnny doesn't want to get eaten by a monster, and has therefore been sleeping with the lights on ever since he can remember. Taking Johnny's axioms as valid assumptions, convince him using a *reductio ad absurdum* that monsters do not eat little boys.

2.2 The definition of the derivative

The single most important application of the limit is that it gives us a way to formalize the idea of a derivative, which we have so far been using on an informal basis. We start from the Newton-Leibniz approach described on p. 47, but modify it by using a limit to get rid of the questionable procedure of discarding the square of an infinitesimal number.

Definition of the derivative

The derivative of a function f at a point x is

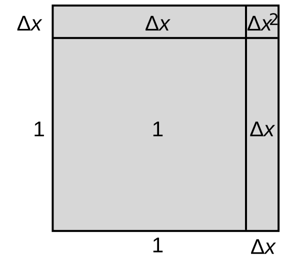
$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

The derivative of x^2, using limits Example 5

Let's use the definition to find the derivative of x^2 at $x = 1$. We have

$$\begin{aligned} f'(1) &= \lim_{\Delta x \to 0} \frac{(1+\Delta x)^2 - 1}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{2\Delta x + \Delta x^2}{\Delta x} \\ &= \lim_{\Delta x \to 0} (2 + \Delta x) \end{aligned}$$

We've already shown in example 4 on p. 50 that this sort of limit of a linear function is just what you would expect by plugging in to the equation of the line, and therefore we have $f'(1) = 2$.



b / A geometrical interpretation of the expression $2\Delta x + \Delta x^2$ occurring in the second line of example 5. The area gained by increasing the size of the square equals the area of the two thin strips plus the area of the small square.

The derivative of an exponential function, with limits Example 6

In example 3 on p. 19, we inferred using a simple geometrical trick that the derivative of an exponential function like $f(x) = 2^x$ must be proportional to f itself,

$$f' = kf,$$

where the constant of proportionality k depends on the base, such as 2. We can now prove the same fact using limits, and say something about the value of the constant. Since this fact is supposed to hold for all values of x, and k is to be the same for any x, we can pick any convenient value for x, say $x = 0$. For the derivative we have

$$\begin{aligned} f'(0) &= \lim_{\Delta x \to 0} \frac{2^{0+\Delta x} - 2^0}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{2^{\Delta x} - 1}{\Delta x}. \end{aligned}$$

Since $f(0) = 1$, we have

$$k = \lim_{\Delta x \to 0} \frac{2^{\Delta x} - 1}{\Delta x}$$

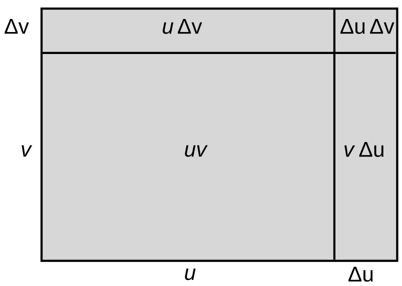
We can get as good an approximation to this limit as we like by plugging in small enough values of Δx. For example, $\Delta x = 10^{-4}$ gives $k \approx 0.69317$, which seems to be an approximation to $\ln 2 = 0.69314\ldots$ This naturally leads us to conjecture that the derivative of b^x equals $(\ln b)b^x$, and in particular that the derivative of e^x is simply e^x. This is investigated further in section 5.2, p. 126.

If the limit referred to in the definition of the derivative is undefined at a certain x, then the derivative is undefined there, and we say that f is not *differentiable* at x. Differentiability is discussed in more detail in section 2.8, p. 61.

We seldom evaluate a derivative by directly applying its definition as a limit. Instead, we use a variety of other more convenient rules that follow from the definition. Some of these are the properties in section 1.2.3, p. 16. In addition, we will learn two very important and useful rules, the product rule and the chain rule.

2.3 The product rule

The idea behind the product rule is very similar to the geometrical intuition expressed by figure b on p. 51 for the derivative of x^2. Suppose that instead of x multiplied by x to make x^2, we have some other function such as $(x^2+7)(x^3)$, which is also the product of two factors. Call these factors $u(x)$ and $v(x)$, so that the function we're differentiating is $f(x) = u(x)v(x)$. Then the expression we get by applying the definition of the derivative to f can be written in terms of the rectangular areas in figure c as



c / A geometrical interpretation of the product rule.

$$f'(x) = \lim_{\Delta x \to 0} \frac{(\text{right strip}) + (\text{top strip}) + (\text{tiny box})}{\Delta x}$$

One can prove from the definition of the limit that the limit of a sum is equal to the sum of the limits, provided that the individual limits exist (see section 4.1, p. 95, property P_3), so:

$$\begin{aligned} f'(x) = &\lim_{\Delta x \to 0} \frac{(\text{right strip})}{\Delta x} \\ &+ \lim_{\Delta x \to 0} \frac{(\text{top strip})}{\Delta x} \\ &+ \lim_{\Delta x \to 0} \frac{(\text{tiny box})}{\Delta x} \end{aligned}$$

If the functions u and v are both well-behaved at x (specifically, if both of them are differentiable), then the "tiny box" term will vanish upon application of the limit just as in example 5. We then have

$$\begin{aligned} f'(x) &= \lim_{\Delta x \to 0} \frac{(\text{right strip})}{\Delta x} + \lim_{\Delta x \to 0} \frac{(\text{top strip})}{\Delta x} \\ &= u'(x)v(x) + v'(x)u(x). \end{aligned}$$

We have the following extremely important and useful rule for differentiation:

Product rule

Let $f = uv$, where f, u, and v are all functions. Then at any point where u and v are both differentiable,

$$f' = u'v + v'u.$$

The product rule for x^3 Example 7

So far we have never actually proved any derivatives of powers of x other than x^2; although the proofs can be done by the methods of ch. 1, they are tedious. These results come out much more easily by applying the product rule. We have already proved that the derivative of x^2 was $2x$. To get the derivative of x^3, we can simply rewrite it as the product $(x^2)\cdot(x)$. Applying the product rule then gives

$$\begin{aligned}(x^3)' &= [(x^2)\cdot(x)]' \\ &= (x^2)'\cdot(x) + (x^2)\cdot(x)' \\ &= 2x\cdot x + x^2\cdot 1 \\ &= 3x^2.\end{aligned}$$

A dirty trick for finding the derivative of $1/x$ — Example 8

How do we differentiate $1/x$? We can guess the right result by recalling that this expression can also be written as x^{-1}. (Exponents, including negative ones, will be reviewed more systematically in section 2.5, p. 56). If we then assume that the power rule $(x^n)' = nx^{n-1}$ applies to $n=-1$, then the result should be that the derivative of $1/x$ is $-x^{-2}$, or $-1/x^2$.

But that's only a reasonable guess, not a proof. We can prove it by the following dirty trick. Write $1 = (x)(1/x)$, and then differentiate on both sides. The left-hand side is a constant, so its derivative is zero. Applying the product rule to the right-hand side, we get $(x)'(1/x) + (x)(1/x)'$, and equating this to zero shows that indeed, $(1/x)' = -1/x^2$.

2.4 The chain rule

2.4.1 Constant rates of change

In addition to the product rule, the other extremely important rule for differentiation is the chain rule. We start with three examples that illustrate the idea but don't require calculus.

Burning calories — Example 9

▷ Jane hikes 3 kilometers in an hour, and hiking burns 70 calories[2] per kilometer. At what rate does she burn calories?

▷ We let x be the number of hours she's spent hiking so far, y the distance covered, and z the calories spent. Then

$$\begin{aligned}\frac{\Delta z}{\Delta x} &= \frac{\Delta z}{\Delta y}\cdot\frac{\Delta y}{\Delta x} \\ &= \left(\frac{70\ \text{cal}}{1\ \cancel{\text{km}}}\right)\left(\frac{3\ \cancel{\text{km}}}{1\ \text{hr}}\right) \\ &= 210\ \text{cal/hr}.\end{aligned}$$

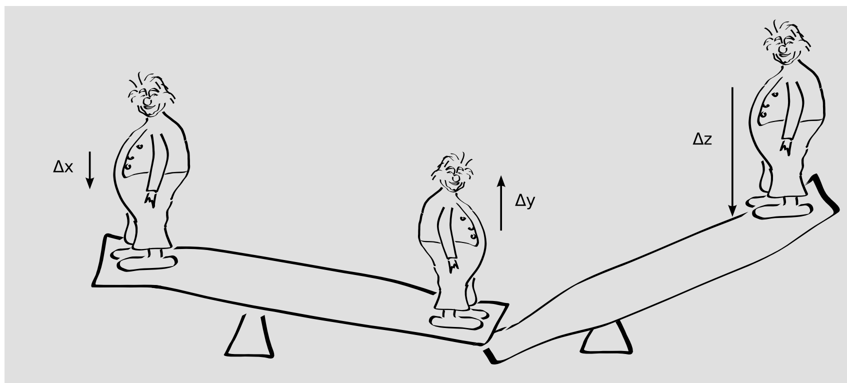
Clowns on seesaws — Example 10

In figure d, the clown on the left drops by Δx, causing the middle clown to go up by Δy. The ratio between these appears to

[2] Food calories are actually *kilo*calories, 1 kcal=1000 cal.

be about $-3/2$ based on the lengths of the two lever arms, as determined by the position of the fulcrum. This then causes the right-hand clown to drop by Δz, where $\Delta z/\Delta y$ is about -2. The result is

$$\frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y} \cdot \frac{\Delta y}{\Delta x} = (-2)(-\frac{3}{2}) = 3.$$



d / Example 10.

e / Example 11.

f / The chain rule allows us to differentiate expressions in which functions occur nested inside other functions, like Russian dolls.

Gear ratios Example 11

▷ Figure e shows a piece of farm equipment containing a train of gears with 13, 21, and 42 teeth. If the smallest gear is driven by a motor, relate the rate of rotation of the biggest gear to the rate of rotation of the motor.

▷ Let x, y, and z be the angular positions of the three gears. Then

$$\frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y} \cdot \frac{\Delta y}{\Delta x} = \frac{13}{21} \cdot \frac{21}{42} = \frac{13}{42}.$$

These examples all used the following relationship among three rates of change:

$$\frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y} \cdot \frac{\Delta y}{\Delta x} \tag{1}$$

Because the rates of change were stated to be constant, it was valid to measure them with expressions of the form $\Delta \ldots / \Delta \ldots$, and because the deltas were real numbers, it was valid to use the normal rules of algebra and cancel the factors Δy.

2.4.2 Varying rates of change

The Leibniz notation makes it tempting to simply write down and believe the following analogous-looking expression involving derivatives:

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\mathrm{d}z}{\mathrm{d}y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x}$$

In problems p1-p3 on p. 43 we verified that this seemed to work. But how do we know that this always works with derivatives? If we define the Leibniz notation as standing for a limit, then we need to show this:

$$\lim_{\Delta x \to 0} \frac{\Delta z}{\Delta x} = \left(\lim_{\Delta y \to 0} \frac{\Delta z}{\Delta y} \right) \left(\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \right) \qquad (2)$$

Rather than giving a formal proof, I've briefly sketched in Box 2.2 the technical issues involved. These work out as our intuition suggests, and we therefore have:

The chain rule
If z is a function of y, and y is a function of x, and if the derivatives $\mathrm{d}z/\mathrm{d}y$ and $\mathrm{d}y/\mathrm{d}x$ exist at a certain point, then at that point,

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\mathrm{d}z}{\mathrm{d}y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x}.$$

The chain rule is extremely useful in evaluating derivatives, because many of the expressions we want to differentiate have a structure in which a big formula is built out of smaller ones. For example, in problem r1 on p. 44, we found by numerical approximation that the derivative of the function

$$\frac{1}{1-x},$$

evaluated at $x = 0$, was about 1.000. The chain rule gives us an easy way to get an exact result for any x. The structure of our formula is like this:

$$\boxed{\frac{1}{\boxed{1-x}}}$$

In silly notation, the chain rule says:

$$\frac{\mathrm{d}\boxed{}}{\mathrm{d}x} = \frac{\mathrm{d}\boxed{}}{\mathrm{d}\boxed{}} \cdot \frac{\mathrm{d}\boxed{}}{\mathrm{d}x}$$

▷Box 2.2 **A sketch of the technical issues behind the chain rule**

If all three derivatives in equation (2) exist, then the equation essentially works because the limit of a product is the product of a limit (provided that the limits exist); this is property P_5 of the limit, to be discussed more formally in section 4.1, p. 95. There are two other technical issues to worry about.

First, equation (1) is not true if $\Delta y = 0$, because we can't divide by zero, and if the derivative of y with respect to x happens to be zero somewhere, then it's reasonable to worry that this might be forced upon us for a certain value of Δx. Although we won't prove it here, this issue doesn't actually cause the chain rule to fail.

The second issue is that in equation (2), two of the limits involve $\Delta x \to 0$, but one has $\Delta y \to 0$. This turns out not to be a problem because, as discussed in ch. 4, a differentiable function must be *continuous* (i.e., there are no gaps in its graph), and therefore if, by assumption, y is differentiable as a function of x, then y is also continuous, and therefore taking $\Delta x \to 0$ also causes $\Delta y \to 0$.

Writing the boxes inside the equations is cumbersome, so let's call the big box z and the small one y. Then

$$z = 1/y \qquad \text{and}$$
$$y = 1 - x,$$

which are both functions we know how to differentiate:

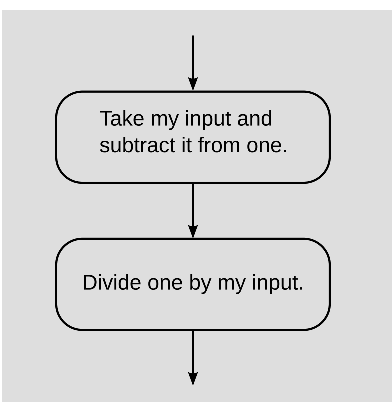
$$\frac{\mathrm{d}z}{\mathrm{d}y} = -y^{-2} \qquad \text{[example 8, p. 53]}$$
$$\frac{\mathrm{d}y}{\mathrm{d}x} = -1$$

In life, sometimes our big goals (get married and raise a family) break down into smaller sub-goals (buy a ring, find a priest, placate the mother of the bride). The chain rule lets us apply this divide-and-conquer strategy to differentiation. Since we know how to differentiate z with respect to y and y with respect to x, the chain rule lets us solve the larger problem of differentiating z with respect to x:

$$\begin{aligned}\frac{\mathrm{d}z}{\mathrm{d}x} &= \frac{\mathrm{d}z}{\mathrm{d}y}\cdot\frac{\mathrm{d}y}{\mathrm{d}x}\\ &= (-y^{-2})(-1)\\ &= y^{-2}\\ &= (1-x)^{-2}.\end{aligned}$$

Plugging in $x = 0$, we verify that the derivative is exactly equal to 1, in agreement with the earlier numerical calculation.

g / Composition of functions is like a bucket brigade. (The workers in the photo are salvaging inventory from a warehouse after the 2010 earthquake in Haiti.)



h / The function $1/(1-x)$ can be viewed as a rule for a two-step computation in which the output of the first computation is fed through as the input to the second stage.

2.4.3 Composition of functions

A little more formally, we can view the chain rule as a rule for doing calculus on functions that are built by *composition* of other functions. The composition $g \circ h$ of functions g and h means the function that takes an input x and gives back an output $g(h(x))$. That is, we take the input x, stick it into h, take h's output, put it in g, and finally take g's output.

The chain rule tells us how to differentiate a function built out of such a composition. In terms of this notation, suppose that $f(x) = g(h(x))$. Then the chain rule says that $f'(x) = g'(h(x))h'(x)$. Or, in a simpler but more abstract notation, we can write $(g \circ h)' = (g' \circ h)h'$.

2.5 Review: exponents that aren't natural numbers

In section 2.6 we will exploit the product and chain rules to prove the rule $(x^n)' = nx^{n-1}$ for all values of n that are nonzero rational numbers. As preparation, we review in this section the basic idea of exponentiation, and then the interpretation of exponents that aren't natural numbers.

2.5.1 Basic ideas

We can represent repeated multiplication

$$2 \times 2 \times 2 = 8,$$

using the notation for exponents,

$$2^3 = 8.$$

Because multiplication is associative,

$$2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 128$$

is the same as

$$(2 \times 2 \times 2)(2 \times 2 \times 2 \times 2),$$

so 2^7 is the same as $(2^3)(2^4)$. In other words, multiplication is the same as adding exponents,

$$b^u b^v = b^{u+v}. \qquad (3)$$

An important special case is scientific notation, which uses powers of 10. For example, $(10^2)(10^7) = 10^9$.

2.5.2 Zero as an exponent

Suppose we compute the list of decreasing powers of a given base, for example $2^3 = 8$, $2^2 = 4$, and $2^1 = 2$. Each result is half as big as the previous one. Therefore if we want to continue reducing the exponent, we should clearly have $2^0 = 1$ in order to continue the pattern. In general, $b^0 = 1$ for any nonzero base b. (The special case 0^0 is undefined.)

2.5.3 Negative exponents

Continuing this pattern, we must have $2^{-1} = 1/2$. In general, negative exponents indicate the inverse of the corresponding positive exponent.

2.5.4 Fractional exponents

Our rules for zero and negative exponents were consistent with equation (3). We can also define fractional exponents that obey this rule. For example, if $3^{1/2}$ is a number, then equation (3) requires that $(3^{1/2})(3^{1/2}) = 3$, so an exponent $1/2$ must mean the same thing as a square root.

2.5.5 Irrational exponents

If we want to define an expression such as 2^π, we can take it to be the limit of the list of numbers 2^3, $2^{3.1}$, $2^{3.14}$, $2^{3.141}$, ...

2.6 Proof of the power rule in general

In section 1.3, p. 20, I presented the rule $(x^n)' = nx^{n-1}$ for all natural numbers n, but only explicitly proved it for $n = 1$ and 2.

A good application of the product and chain rules is to extend the proof to all nonzero integers n and to show that it also holds for fractional exponents.

Only $n = 0$ requires special treatment. Since $x^0 = 1$, its derivative should be zero. Our rule sort of, but not quite, works here, since it gives $0x^{-1}$, or $0/x$. This is certainly zero if $x \neq 0$, but in the case where $x = 0$ it gives $0/0$, which is undefined.

▷Box 2.3 **Ideas about proof: proof by induction**

Proof by induction is a technique for proving an infinite number of facts without using infinitely many words. Call these facts, or *propositions*, P_1, P_2, and so on. For example P_n could be the claim that if we kick over the first in an infinite chain of dominoes, then the nth domino will fall as well. Induction requires two steps.

(1) We establish that P_1 is true. For example, if we kick over the first domino, then P_1 is clearly true, since kicking it over causes it to fall. This is called the *base case*.

(2) We show that if P_{n-1} holds, then P_n is true as well. For example, if domino $n-1$ falls, then it will cause domino n to fall as well.

2.6.1 Exponents that are natural numbers

Example 7 on p. 52 showed that the product rule can be used to prove special cases of the power rule. Since we knew the derivative of x^2, we were able to find the derivative of x^3 by rewriting it as $(x^2)(x)$ and applying the power rule. In the same way, we can prove the rule for any exponent n if it has already been established for $n-1$. We rewrite x^n as $(x^{n-1})(x)$, differentiate using the product rule, and find:

$$\begin{aligned}(x^n)' &= (x^{n-1})'(x) + (x^{n-1})(x)' \\ &= (n-1)x^{n-2}x + x^{n-1} \\ &= nx^{n-1}\end{aligned}$$

By establishing the fact for $n = 1$, and then proving that it must hold for n if it holds for $n-1$, we establish that it holds for all natural numbers n. This is called *proof by induction* (box 2.3).

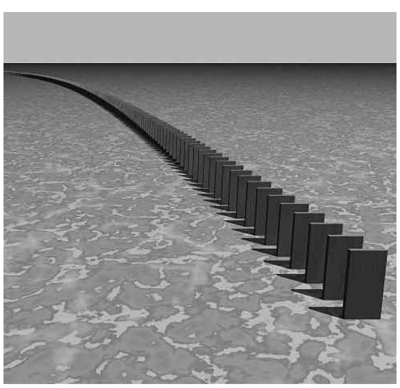
2.6.2 Negative exponents

We saw in example 8 on p. 53 that $(1/x)' = -1/x^2$, which was exactly what we would have expected from applying the power rule to the exponent -1. It is then straightforward to extend the result to all negative integers by applying the chain rule to $(x^n)^{-1}$.

2.6.3 Exponents that aren't integers

What about fractional exponents, such as $x^{1/2}$, i.e., the square root of x? We don't know what this derivative is yet, but let's give it a name. Call it f, i.e., $f(x) = (\sqrt{x})'$. Then

$$\begin{aligned}1 &= x' \\ &= (\sqrt{x}\sqrt{x})' \\ &= f(x)\sqrt{x} + \sqrt{x}f(x) \\ &= 2f(x)\sqrt{x} \\ f(x) &= \frac{1}{2\sqrt{x}} \\ &= \frac{1}{2}x^{-1/2}\end{aligned}$$



i / Proof by induction is like an infinite chain of dominoes. If we topple the first domino, then eventually every domino will fall.

This is exactly what we would have inferred from the power rule $(x^n)' = nx^{n-1}$, with $n = 1/2$. A similar argument can be carried out for any fractional exponent, although recognizing this is not quite the same as writing a general proof; a general proof is given in example 8, p. 165. The generalization to irrational exponents is deferred until example 4 on p. 135.

Economic order quantity — Example 12

Here is an extremely common problem in the business world. A retailer knows that there is a steady yearly demand D for the widgets it sells; every year, customers buy D widgets. They need to maintain an inventory of the product, and when they run out, they need to buy a quantity q from their wholesaler. Ordering from the wholesaler costs a certain amount per widget plus a certain amount per order, and because of the per-order cost, the retailer would prefer that the quantity of widgets q in each order be big.

The retailer also has to pay a certain amount to store all the widgets in inventory. For example, if their inventory gets too big, they may have to buy or rent a new warehouse. This is a reason not to make q too big.

We have the following model of the retailer's yearly costs:

$$\begin{aligned} C = {} & c_1 D && \text{[wholesale cost of the widgets, including shipping]} \\ & + c_2 \frac{D}{q} && \text{[D/q=number of orders; c_2=fixed cost per order]} \\ & + c_3 q && \text{[cost of storing an inventory of q widgets]} \end{aligned}$$

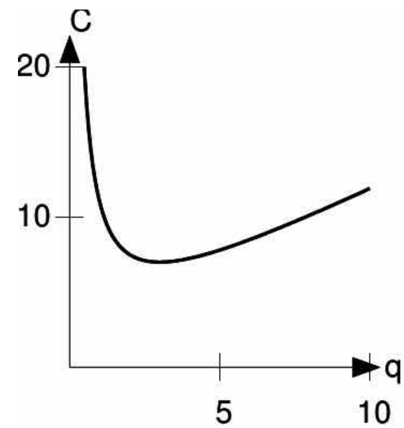
We want to minimize the function $C(q)$, taking D, c_1, c_2, and c_3 as constants. If q is too small, the second term dominates and becomes large, while the same happens with the third term if q is too big. Therefore we know that the minimum of C must occur at some finite value of q. The function is smooth, so this minimum must occur at a point where the derivative $\mathrm{d}C/\mathrm{d}q$ is zero (section 1.5.3, p. 24). Writing $1/q$ as q^{-1} and applying the power rule, the derivative is

$$\frac{\mathrm{d}C}{\mathrm{d}q} = -c_2 D q^{-2} + c_3,$$

and setting this equal to zero gives

$$q = \sqrt{\frac{c_2 D}{c_3}},$$

where only the positive square root has real-world significance. This answer makes sense because we respond to greater demand D by making bigger orders, and likewise if the fixed cost per order c_2 is high, we will make bigger orders in order to reduce the number of orders. If the cost c_3 of warehousing a widget for a year is large (e.g., the widget is a jumbo jet), then we will order in smaller quantities.



j / Example 12, with $c_1 D = 1$, $c_2 D = 9$, and $c_3 = 1$.

2.7 Quotients

Suppose that we want to differentiate the function

$$\frac{1}{x}.$$

The product rule tells us how to differentiate an expression involving multiplication, but this one uses division. However, division by a certain number is the same as multiplication by its multiplicative inverse, so we can rewrite this function in a form that we know how to differentiate.

$$\begin{aligned}\left(\frac{1}{x}\right)' &= \left(x^{-1}\right)' \\ &= -x^{-2} \qquad \text{[power rule]}\end{aligned}$$

If the expression in the denominator is more complicated, we can do the same thing, but use the chain rule as well:

$$\begin{aligned}\left(\frac{1}{1+x^2}\right)' &= \left((1+x^2)^{-1}\right)' \\ &= -(1+x^2)^{-2}(2x)\end{aligned}$$

If the numerator is not just 1, then we also have to use the product rule:

$$\begin{aligned}\left(\frac{x^3}{1+x^2}\right)' &= \left(x^3(1+x^2)^{-1}\right)' \\ &= (x^3)'(1+x^2)^{-1} + x^3\left[(1+x^2)^{-1}\right]' \qquad \text{[product rule]} \\ &= 3x^2(1+x^2)^{-1} + x^3\left[-(1+x^2)^{-2}(2x)\right] \\ &= \frac{x^4+3x^2}{(1+x^2)^2} \qquad \text{[simplify]}\end{aligned}$$

The foregoing examples show a technique for differentiating quotients that works in all cases, and this is how I do that type of derivative. Some people, however, prefer to memorize the following rule, which can be proved by running through the steps above for a function $f = p/q$, where p and q can be any functions at all.

Quotient rule
Let $f = p/q$, where f, p, and q are all functions. Then at any point where p and q are both differentiable and $q \neq 0$,

$$f' = \frac{p'q - q'p}{q^2}.$$

In the examples above, the functions p and q happened to be polynomials. A function like f that is formed in this way from the quotient of polynomials is called a *rational function*.

2.8 Continuity and differentiability

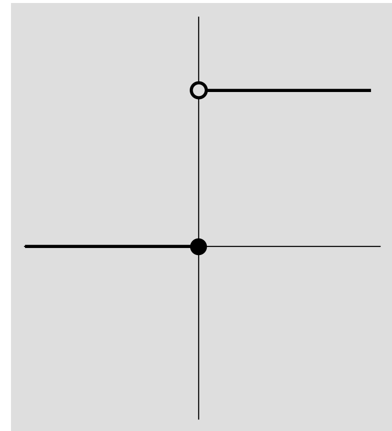
2.8.1 Continuity

Intuitively, a continuous function is one whose graph has no sudden jumps in it; the graph is all a single connected piece. Such a function can be drawn without picking the pen up off of the paper. Formally, continuity is defined as follows.

A function g is *continuous* at a if

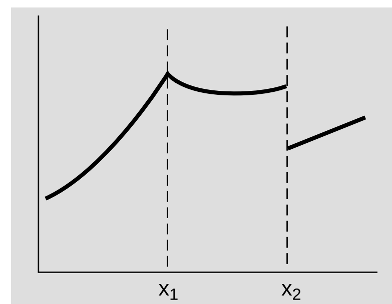
$$\lim_{x\to a} g(x) = g(a) \tag{4}$$

A function is continuous if it is continuous at every a in its domain.



k / A discontinuous function.

In most cases, there is no need to invoke the definition explicitly in order to check whether a function is continuous. Most of the functions we work with are defined by putting together simpler functions as building blocks. For example, let's say we're already convinced that the functions defined by $g(x) = 3x$ and $h(x) = \sin x$ are both continuous.[3] Then if we encounter the function $f(x) = \sin(3x)$, we can tell that it's continuous because its definition corresponds to $f(x) = h(g(x))$. The composition of two continuous functions is also continuous. Just watch out for division. The function $f(x) = 1/x$ is continuous everywhere except at $x = 0$, so for example $1/\sin(x)$ is continuous everywhere except at multiples of π, where the sine has zeroes.



l / The function is not differentiable at x_1 because it has a kink there, and is not differentiable at x_2 because it has a sudden jump.

2.8.2 More about differentiability

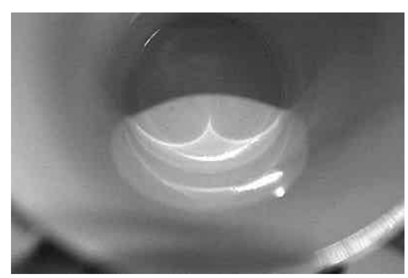
We mentioned briefly on p. 51 that a function is defined to be differentiable or nondifferentiable at a particular point depending on the existence of the limit referred to in the definition of the derivative,

$$f'(x) = \lim_{\Delta x\to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

Figure l shows two common reasons why a function would not be differentiable at a certain point: because it has a kink, or because it is discontinuous. If a function is discontinuous at a given point, then it is not differentiable at that point.

Although differentiability implies continuity, a function can be continuous without being differentiable; see example 13.

We seldom have to resort to limits and epsilon-delta arguments in order to determine whether a function is differentiable at a particular point. Here are three methods that, when they apply, are usually easier:



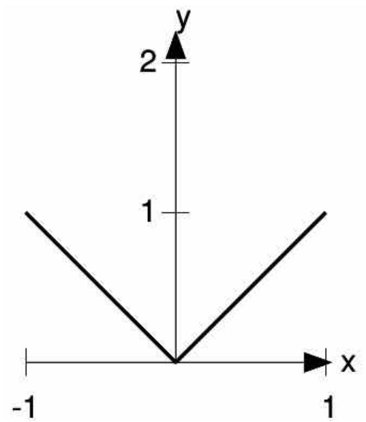
m / Reflected light forms a geometrical curve inside a teacup. The curve has a kink similar to the one at x_1 in figure l. This kink is of a special type called a *cusp*, in which the two branches are parallel where they meet.

[3] The reader who has forgotten all of his/her trig is directed to the review in section 5.3.

1. Graph the function and apply the informal definition of the derivative from section 1.2.1, p. 14. That is, imagine trying to zoom in on the point of interest until the curve appears straight, and then measuring its slope. If something goes wrong in this process, then the function isn't differentiable.

2. Often we deal with functions that have been defined by a formula, which means building it out of other functions through arithmetic operations and composition. If all of these functions and operations are differentiable at the point of interest, then the function is differentiable.

3. If the function f has been defined by a formula, then it will usually be possible differentiate it using the differentiation rules and write the result as a new formula for f'. Often there will be only certain specific points where the formula for f' is undefined, so these are the points where f wasn't differentiable.

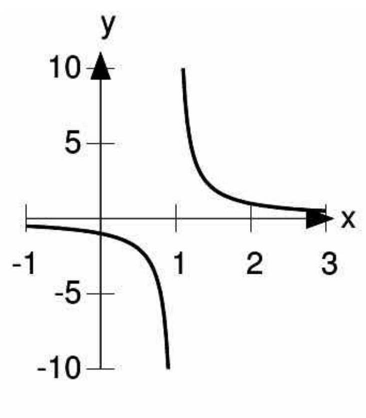


n / Example 13.

The absolute value function Example 13

▷ Where is the function $y = |x|$ differentiable?

▷ By visualizing the graph, figure n, and applying method 1 we can tell immediately that it's differentiable everywhere except at $x = 0$. At $x = 0$, there is a kink, and no matter how far we zoom in, the kink will never look like a line.

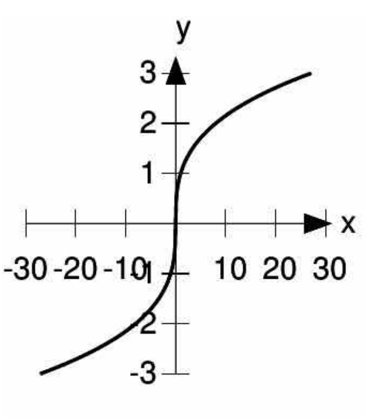


o / Example 14.

Not differentiable when dividing by zero Example 14

▷ Where is the function $f(x) = 1/(x-1)$ differentiable?

▷ Let's use method 2 above. This function can be built out of the composition of functions as $f(x) = g(h(x))$, where $g(x) = 1/x$ and $h(x) = x - 1$. Both of these functions are well-behaved everywhere, except that g isn't differentiable where it blows up at $x = 0$. Therefore the function f is differentiable everywhere except at $x = 1$, which is where $h(x) = 0$ is the input to $g(x)$.



p / Example 15.

Differentiability of the cube root Example 15

▷ Where is the function $y = x^{1/3}$ differentiable?

▷ Let's use method 3. The power rule gives $y' = \frac{1}{3}x^{-2/3}$. This is well defined everywhere except at $x = 0$, where it blows up to infinity. Therefore y is differentiable everywhere except at $x = 0$.

Nondifferentiable ingredients, differentiable result Example 16

Method 2 can prove that a function is differentiable, but cannot necessarily be used to prove it nondifferentiable. For example, consider the function $y = x^5(1 + 1/x)$. The second factor blows up to infinity at $x = 0$, which makes us suspect that y is not differentiable there. But in fact the formula can be rewritten as $y = x^5 + x^4$, which is clearly differentiable everywhere. Although the second factor in the original form blows up at $x = 0$, the first factor vanishes there so rapidly that the product also vanishes, and vanishes smoothly.

2.8.3 Zero derivative at the extremum of a differentiable function

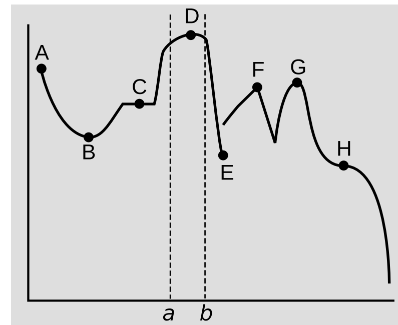
We saw in section 1.5.3, p. 24, that although a searching for a zero derivative may be a good way to find an extremum, it doesn't always work. Looking at the zoo of possibilities in figure q, we see that both of the following statements are false:

1. If a function has a local extremum, it must have a zero derivative there. (False: fails at A, E, and F.)
2. If a function has a zero derivative somewhere, that must be a local extremum. (False: fails at H.)

In mathematical jargon, we say that a zero derivative is neither a necessary (1) nor a sufficient (2) condition for a local extremum.

We can, however, make a more restricted statement of 1 that is true.

Theorem
If a function f is continuous on an interval $[a, b]$ and differentiable on (a, b), and if there is a point $c \in (a, b)$ for which $f(c)$ is a maximum or minimum in the interval, then $f'(c) = 0$.



q / If point D is a maximum over the interval $[a, b]$, then f' equals zero at D.

Let's see why all the conditions are necessary. The assumption of continuity is needed because of points like E. We need differentiability because of F. We also needed to assume that c was on the interior of the interval, since otherwise it would have been possible to choose b so that point E lay at $x = b$.

Proof: We prove the case where $f(c)$ is a maximum, as in figure q; the other case is exactly analogous. Since f is assumed to be differentiable, it's differentiable at c, and since c is on the interior of the interval, differentiability means that the derivative must have the same value regardless of whether we approach c from the right or from the left. (At a nondifferentiable point such as F, the two limits could be unequal.) Let's look at both of these limits. The limit from the left is

$$\lim_{h \nearrow 0} \frac{f(c+h) - f(c)}{h}.$$

But since we assumed $f(c)$ to be the greatest value on $[a, b]$, the quantity inside the limit is guaranteed to be greater than or equal to zero. The limit exists, since we assumed differentiability, so the limit must also be greater than or equal to zero. Similarly, the limit from the right

$$\lim_{h \searrow 0} \frac{f(c+h) - f(c)}{h}$$

must exist and be *less than* or equal to zero. Since the two limits are equal, they equal zero.□

2.9 Safe handling of d*y* and d*x*

We've seen that although the real number system doesn't include infinitely big or infinitely small quantities, it can nevertheless be extremely useful to think of a notation like $\mathrm{d}y/\,\mathrm{d}x$ as the quotient of two infinitely small numbers. For example, it allows us to check our work in differentiation by checking the units of the result (example 8, p. 28), and it makes the chain rule look so obvious that there would never be any danger of forgetting it. When the calculus was first invented, these infinitely small numbers were referred to as *infinitesimal* numbers. The idea behind the word is that just as a decimal is one tenth, an infinitesimal is one "infinitieth."

We now confront the question of when it's safe to treat $\mathrm{d}y$ and $\mathrm{d}x$ as if they were numbers. This kind of manipulation is like nuclear energy: it can be used for good and for evil, and if you want to use it safely, you have to know what you're doing. In this section we lay out some simple safety rules which, if followed, will prevent all nuclear meltdowns. Just as we enriched the set of natural numbers to make the rational numbers, and the rational numbers to make the reals, we continue the march of progress by making an even larger number system called the hyperreal numbers, which includes infinitesimals. For a more detailed exposition at the freshman-calculus level, see the excellent free online book by Keisler, *Elementary Calculus: An Approach Using Infinitesimals*.

We start with two preliminary definitions.

Definition: Suppose that for a certain nonzero number d, we have $|d| < 1$, $|d| < 1/(1+1)$, $|d| < 1/(1+1+1)$, ... and so on for all inequalities of this form.[4] Then we say that d is *infinitesimal*.

Definition: Let H be a hyperreal number (which may or may not also be a real number). Suppose that there exists some real number r such that $|H - r|$ is infinitesimal. Then we say that r is the *standard part* of H.

Rule 1. The hyperreal numbers obey all the same elementary axioms as the real numbers (section 1.6, p. 25).

The hyperreals numbers include at least one infinitesimal number, call it d. By rule 1, we can apply the multiplicative inverse axiom to d, so $1/d$ is also a well-defined hyperreal number, and clearly $1/d$ is bigger than 1, bigger than $1 + 1$, and so on, so the hyperreal number system includes both infinitely big and infinitely small quantities.

[4] Cf. example 11, p. 113. For an application to economics, see rule 3, p. 218.

It can be proved from the elementary axioms that if d is nonzero, then $2d \neq d$. Therefore the hyperreal number system includes a variety of sizes of infinitesimals. This is important, because if all infinitesimals were the same size, then $\mathrm{d}y/\,\mathrm{d}x$ would always have to equal one! It also follows from the axioms that $1/d \neq 1/(2d)$, so infinite numbers come in different sizes as well. We therefore have:

Rule 2. The symbol ∞ and the term "infinity" do not stand for any real number, and do not stand for any specific hyperreal number. They are in fact not very useful in the context of the hyperreals.

Breaking the rules gives a nuclear meltdown Example 17

Suppose that the universe is infinite, so that there are infinitely many animals in the universe that, like us, have two eyes. The number of left eyes is some infinite hyperreal number H, and H is also the number of right eyes. The total number of eyes is then

$$H + H = 2H.$$

Everything is all right, and $2H$ is an infinite number that happens to be twice as big as H.

But now suppose we break rule 2 and use the symbol ∞ indiscriminately for any positive, infinite quantity. Then we have

$$\infty + \infty = \infty.$$

Applying the additive inverse axiom, we can cancel an ∞ from each side, giving

$$\infty = 0,$$

which is absurd.

The paradox didn't result from talking about infinite numbers. It came from breaking one of the rules for manipulating them correctly.

Historically, one of the main sources of confusion about infinitesimals was the sketchy practice of discarding the square of an infinitesimal (p. 47). This is resolved as follows:

Rule 3. The derivative of y with respect to x is defined as the standard part of $\mathrm{d}y/\,\mathrm{d}x$.

Redoing the example from p. 47 according to this rule, we have the following calculation of the derivative of $y = x^2$ at $x = 1$:

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{(1+dx)^2 - 1}{(1+dx) - 1} \\ &= 2 + dx \\ y' &= \text{standard part of } 2 + dx \\ &= 2 \end{aligned}$$

Although this particular modern approach to calculus makes $\mathrm{d}y/\,\mathrm{d}x$ not a synonym for y', the notational distinction is not assumed in a general context, since they were thought of as synonyms for hundreds of years.

Ideas very much like rules 1 and 3 were in fact originally proposed by Leibniz,[5] but not until the 1960s were they restated precisely enough to satisfy the mathematical community. In the interim, there was considerable suspicion of infinitesimals (Georg Cantor famously referred to them as "infect[ing] mathematics" like a "cholera-bacillus"), and today many mathematicians dislike them, despite their logical rehabilitation, as a matter of taste.

A not-quite proof of the chain rule — Example 18

The Leibniz notation for the chain rule

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

makes it look as though its proof were a matter of trivial algebra: just cancel the factors of dy. This isn't quite valid, however, as a rigorous proof, because the derivative is really not the quotient of two infinitesimals but the *standard part* of that quotient.

▷Box 2.4 **Why 0! equals 1**

We define $0! = 1$, both because it turns out to be more convenient in all of our applications, and for the following logical reason.

In the more usual case where $n \geq 1$, $n!$ is defined as a product containing n factors. If we start with a rubber band, then stretch it successively by all of these factors, we end up stretching it by a factor of $n!$ over all.

In the case of $n = 0$, we have no factors in our list, so we have nothing on our list of things to do to the rubber band. It is left at its original length. It has been stretched by a factor of 1, i.e., left alone.

Note that exactly the same logic applies to exponents, and that's why we also define, for example, $7^0 = 1$.

A calculator for infinite and infinitesimal numbers — Example 19

A web-based calculator at `lightandmatter.com/calc/inf` lets you play with infinite and infinitesimal numbers. It provides one built-in infinitesimal number *d* that satisfies the definition on p. 64. The following example shows some sample calculations.

```
2+2
  4
d+d
  2d
d<1/1000
  true
d>0
  true
```

2.10 The factorial

In a number of places in this course, it will be helpful to know about a function called the *factorial*. The factorial of n, notated $n!$, is defined as the product of all the integers from 1 to n,

$$n! = 1 \cdot 2 \ldots n.$$

For example, 3!, read as "three factorial," is $1 \cdot 2 \cdot 3 = 6$. As a special case, we define 0! to be 1 (not zero), for the reasons given in Box 2.4.

[5] Blaszczyk, Katz, and Sherry, "Ten misconceptions from the history of analysis and their debunking," `arxiv.org/abs/1202.4153`.

2.11 Style

Style is important. If you say true things in poor style, people will decide that you're stupid and ignore you. You know enough calculus to appreciate some examples.

1. Use equals signs. State what it is that you're calculating.

wrong	*right*
$3x(x+4)$	$[3x(x+4)]'$
$3x^2+12x$	$= [3x^2+12x]'$
$6x+12$	$= 6x+12$

2. The Leibniz notations d and $\mathrm{d}/\,\mathrm{d}x$ are operations (like $\sqrt{}$), not numbers.

Question: Differentiate x^2.	Question: Differentiate x^2.
Wrong answer: $\mathrm{d}' = 2x$	Right answer: $\mathrm{d}(x^2)/\,\mathrm{d}x = 2x$
Wrong answer: $\mathrm{d}/\,\mathrm{d}x = 2x$	Right answer: $\mathrm{d}(\ldots)/\,\mathrm{d}x = 2x$

3. Immediately make obvious simplifications.

wrong

$(x^2+3)'$
$= 2x^1+0$

right

$(x^2+3)'$
$= 2x^1+0$ [or don't write this at all]
$= 2x$

4. Simplification should usually reduce the number of symbols.

wrong

$[(x^2+1)^3]'$
$= 3(x^2+1)^2(2x)$
$= 3(x^4+2x^2+1)(2x)$ [uglification]

right

$[(x^2+1)^3]'$
$= 3(x^2+1)^2(2x)$
$= 6x(x^2+1)^2$ [simplification]

wrong

$[1/\sqrt{1+x}]'$
$= [(1+x)^{-1/2}]'$
$= -\frac{1}{2}(1+x)^{-3/2}$
$= \frac{-1}{2(1+x)\sqrt{1+x}}$ [uglification]

right

$[1/\sqrt{1+x}]'$
$= [(1+x)^{-1/2}]'$
$= -\frac{1}{2}(1+x)^{-3/2}$ [Stop here.]

5. Don't use a complicated technique when a simple one will do.

wrong

x'
$= (x^1)'$
$= (1)x^0$
$= 1$

right

$x' = 1$ [known fact]

wrong

$\left(\frac{1}{x^2+1}\right)'$
$= \frac{(1)'(x^2+1)-(1)(x^2+1)'}{(x^2+1)^2}$ [quotient rule]
$= \frac{(0)(x^2+1)-(2x)}{(x^2+1)^2}$
$= -\frac{2x}{(x^2+1)^2}$

right

$\left(\frac{1}{x^2+1}\right)'$
$= [(x^2+1)^{-1}]'$
$= -2x(x^2+1)^{-2}$ [power and chain rules]

Review problems

a1 Compute $3^{2014}3^{-2011}$. ✓

a2 Compare $u = 10^{-10^{10}}$ with $v = 10^{-10^{-10}}$. (Note that exponentiation is not associative, and an expression of the form a^{b^c} is interpreted as $a^{(b^c)}$.

a3 Solve $16^x = 1/2$ for x. ✓

Problems

Example 2 on p. 48 demonstrates a way of guessing a limit by plugging in numbers and making a table of values. Do the same thing in problems b1-b3.

b1

$$\lim_{x \to \pi} (\pi - x) \tan \frac{x}{2}$$

b2

$$\lim_{x \to 0} \frac{|x|}{\sqrt{1 - \cos x}}$$

(As always in this course, trig functions are assumed to take angles in radians. Put your calculator in radian mode.)

b3

$$\lim_{x \to 0} x^{-10} e^{-1/|x|}$$

In example 5 on p. 51 we found the derivative of the function $y(x) = x^2$ *by directly applying the definition of the derivative as a limit. In problems c1-c4, apply the same brute-force technique to the given functions.*

c1 $u(a) = a^3$ at $a = 1$

c2 $p(j) = \frac{1}{j}$ at $j = 1$

c3 $t(c) = \frac{1}{c^2}$ at $c = 1$

c4 $s(n) = \frac{1}{1+n}$ at $n = 1$

e1 Differentiate $\sqrt[3]{x}$ with respect to x. ▷ Solution, p. 227

e2 Differentiate the following with respect to x:
(a) $y = \sqrt{x^2 + 1}$
(b) $y = \sqrt{x^2 + a^2}$
(c) $y = 1/\sqrt{a + x}$
(d) $y = a/\sqrt{a - x^2}$

[Thompson, 1919] ▷ Solution, p. 227

e3 The following table shows the barometric pressure P and average July temperature T for the summit of Mount Everest and the city of Wenzhou, China, which is at the same latitude.

	pressure (kPa)	temperature (°C)
Wenzhou	101	+29
Everest	38	−16

A physical model predicts the following relationship between these two variables:

$$T = T_o + cP^{2/7}$$

Here c is a constant and $T_o = -273\,°\mathrm{C}$ is a constant that converts from degrees Celsius to a temperature scale based on absolute zero.
(a) Estimate c from the data at Wenzhou. ✓
(b) T is a complicated nonlinear function of P, and for some purposes, such as mental estimation, a linear approximation might be more convenient to work with. Find the equation of the tangent line to this function at the point representing the conditions at Wenzhou, and use this equation to calculate the expected temperature at the summit of Everest. This is quite a long extrapolation. How good an approximation is it? ✓

e4 Use the product rule to prove the vertical stretch property of the derivative (p. 16). ▷ Solution, p. 227

In problems g1 and g2, compute each derivative by two different methods: (a) by multiplying out the given expression and then differentiating, and (b) by using the product rule. Make sure that you get the same answer by both methods.

g1 $y = (x^2 + x + 1)\sqrt{x}$. ✓

g2 $y = (x + 5)(x^3 + 1)$. ✓

i1 (a) Consider the function $f(x) = xe^x$, where e is the base of natural logarithms. Use the technique described in section 1.8.1, p. 30, to find $f'(1)$, to three decimal places of precision. $\checkmark$
(b) In example 6, p. 51, we conjectured that the derivative of e^x was simply e^x. This is discussed in greater detail in ch. 5, but for now let's just assume that it's true. Given this fact, use the product rule to differentiate the function f. Check that the result is consistent with your answer to part a. $\checkmark$

i2 We've established the power rule using limits, which are the most common modern tool for defining derivatives. By this rule, the derivative of x^3 is $3x^2$, and evaluating this at $x = 1$ gives a derivative of 3.

Chapter 2 began by showing a more old-fashioned technique for differentiating x^2 at $x = 1$ (p. 47). Apply this technique to x^3 at $x = 1$, and show that it agrees with the result found above.
▷ Solution, p. 228

i3 Differentiate $(2x + 3)^{100}$ with respect to x.
▷ Solution, p. 228

i4 Differentiate $(x + 1)^{100}(x + 2)^{200}$ with respect to x.
▷ Solution, p. 228

i5 Use the chain rule to differentiate $((x^2)^2)^2$, and show that you get the same result you would have obtained by differentiating x^8. [M. Livshits] ▷ Solution, p. 228

i6 In section 2.4.3 on p. 56, we expressed the chain rule without the Leibniz notation, writing a function f defined by $f(x) = g(h(x))$. Suppose that you're trying to remember the rule, and two of the possibilities that come to mind are $f'(x) = g'(h(x))$ and $f'(x) = g'(h(x))h(x)$. Show that neither of these can possibly be right, by considering the case where x has units. You may find it helpful to convert both expressions back into the Leibniz notation.
▷ Solution, p. 228

Compute the derivative of each of the functions in problems j1 and j2 by two different methods: (a) by multiplying out the given expression and then differentiating, and (b) by using the chain rule. Make sure that you get the same answer by both methods.

j1 $y = (1 + x^2)^4$ ✓

j2 $y = (x^2 + x + 1)^2$ ✓

In problems k1-k7, differentiate the given function, and try to simplify your answer as much as possible.

k1 $c(d) = d + 1 + (d + 1)^2$ ✓

k2

$$a(b) = \frac{b - 2}{b^4 + 1}$$

✓

k3

$$g(u) = \left(\frac{1}{1 + u}\right)^{-1}$$

✓

k4 $h(z) = \sqrt{1 - z^2}$ ✓

k5

$$h(t) = \frac{at + b}{ct + d} \qquad (a,\ b,\ c, \text{ and } d \text{ are constants.})$$

✓

k6

$$p(c) = \frac{1}{(1 + c^2)^2}$$

✓

k7

$$s(m) = \frac{m}{1 + \sqrt{m}}$$

✓

In problems m1-m1, j, k, ℓ, and m, are constants. Calculate the given derivatives. Simplify answers where possible.

m1 $\dfrac{\mathrm{d}}{\mathrm{d}s}\left[(\ell s^j + ks)^m\right]$ (where $j \neq 0$ and $m \neq 0$) $\checkmark$

m2 $\dfrac{\mathrm{d}}{\mathrm{d}v}\left(\dfrac{v}{jv+k}\right)$ $\checkmark$

m3 $\dfrac{\mathrm{d}}{\mathrm{d}w}\left[(\ell w + m)\sqrt{jw+k}\right]$ $\checkmark$

m4 $\dfrac{\mathrm{d}}{\mathrm{d}\theta}\left(\dfrac{\ell}{\theta^2-1}\right)$ $\checkmark$

n1 Suppose that we put a stick on a table and use a ruler to measure its length L. According to Einstein's theory of special relativity, if the stick is instead in motion at speed v relative to the ruler, then we get a different, shorter length given by

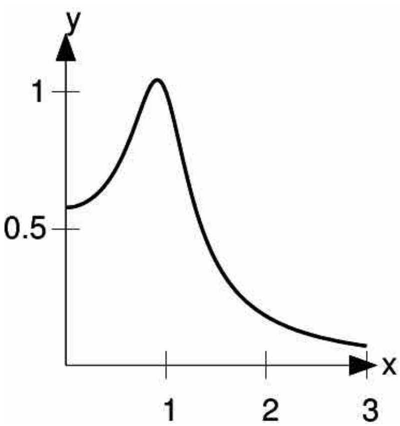
$$M = L\sqrt{1-\frac{v^2}{c^2}},$$

where c is the speed of light. We don't notice this effect in everyday life because ordinary velocities are so small compared to c. (a) Calculate $\mathrm{d}M/\,\mathrm{d}v$, the rate at which the stick shortens with increasing speed. (b) Check the units of your answer (section 1.9, p. 34). (c) Check that the sign of the result makes sense. (d) Discuss the behavior of your result if $v = c$. $\checkmark$

n2 Suppose that a distant galaxy is moving away from us at some fraction u of the speed of light. Then the vibration of the light waves we receive from it is slowed down by the factor

$$D(u) = \sqrt{\frac{1-u}{1+u}}$$

compared to what we would have observed if it hadn't been in motion relative to us. This is called the Doppler effect. Compute the derivative $\mathrm{d}D/\,\mathrm{d}u$, which measures how sensitive the effect is to the velocity. ✓



The function of problem p1, with a = 3, b = 1, and f_o = 1.

p1 When you tune in a radio station using an old-fashioned rotating dial you don't have to be exactly tuned in to the right frequency in order to get the station. If you did, the tuning would be infinitely sensitive, and you'd never be able to receive any signal at all! Instead, the tuning has a certain amount of "slop" intentionally designed into it. The strength of the received signal s can be expressed in terms of the dial's setting f by a function of the form

$$s = \frac{1}{\sqrt{a(f^2 - f_o^2)^2 + bf^2}},$$

where a, b, and f_o are constants. The constant b relates to the amount of slop. This functional form is in fact very general, and is encountered in many other physical contexts. The graph shows an example of the kind of bell-shaped that results curve. Find the frequency f at which the maximum response occurs, and show that if b is small, the maximum occurs close to, but not exactly at, f_o.

▷ Solution, p. 229

p2 Many cactuses are approximately cylindrical in shape. In order to minimize the loss of water through evaporation, it is advantageous for a cactus to have a minimum surface area for a given volume. Find the proportion of height to diameter that achieves this, taking the cactus to be a cylinder with only its top and sides exposed. ✓

p3 An atomic nucleus is made out of protons and neutrons. The number of protons is called Z and the number of neutrons N. Figure s on p. 76 shows a chart of all of the nuclei that have been observed and studied to date. Most of these are unstable: they undergo radioactive decay in a certain amount of time, and therefore are not found in the earth's crust, so they can only be produced artificially.

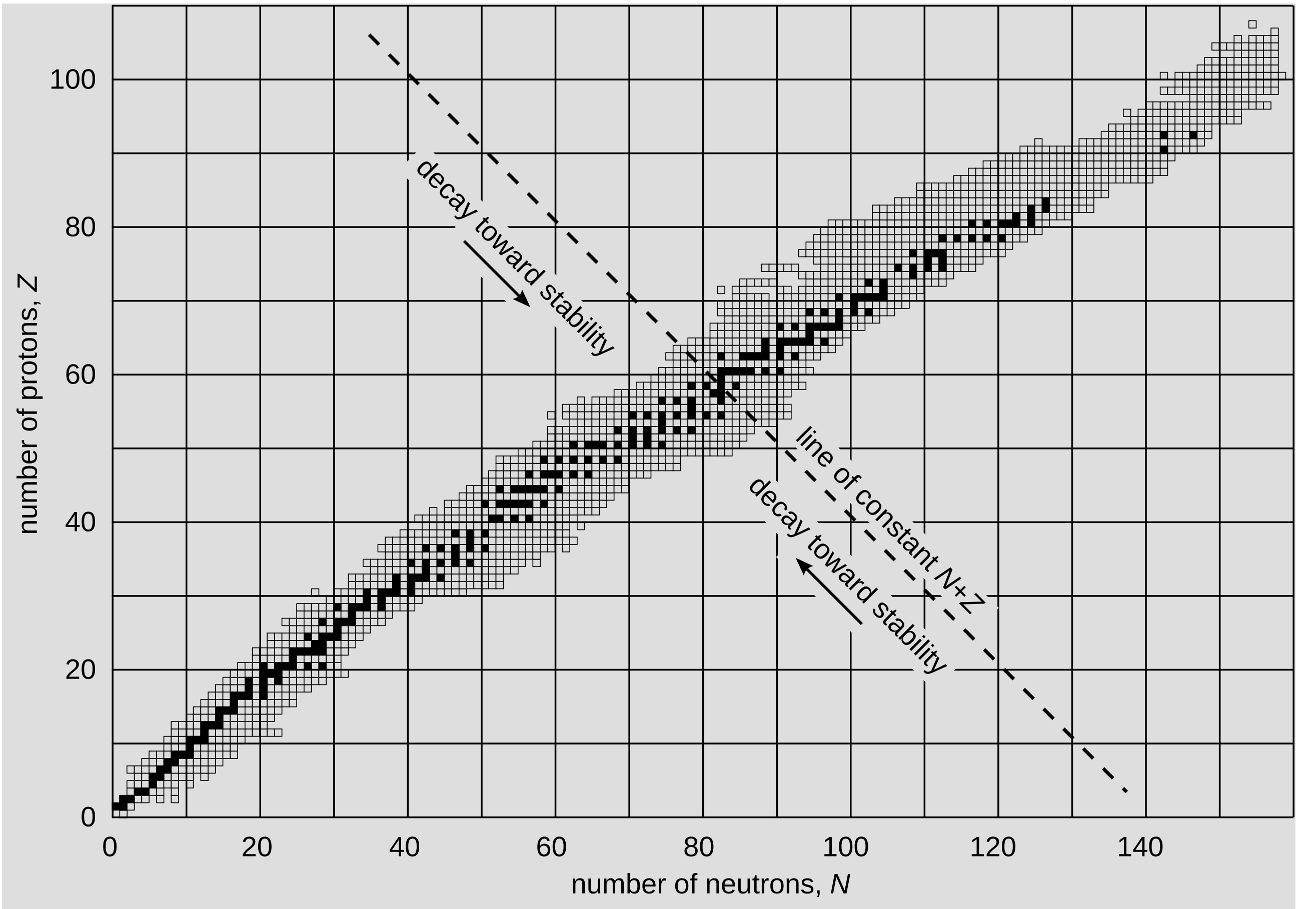
The stable nuclei are shown on the chart as black squares, and we can see that they follow a certain curve. Unstable nuclei that lie below and to the right of the line of stability have too many neutrons in proportion to their protons, and they undergo a decay process in which a neutron is converted to a proton, causing the nucleus to move one step diagonally on the chart, as in the game of checkers. Similarly, nuclei with too few neutrons move by diagonal steps down and to the right. Defining $A = N + Z$, these decay processes keep A constant.

In the *liquid drop model*, the nucleus is treated as a continuous fluid with certain properties such as surface tension. Since the fluid is continuous, we can pretend that N and Z are capable of taking on any real-number values. (This is similar to the water molecules in the reservoir on p. 14.) In this model, a nucleus has a certain energy,

$$E = bZ^2A^{-1/3} + \frac{(A-2Z)^2}{A},$$

where $b \approx 0.031$, and for simplicity we have left out an over-all constant of proportionality with units of energy. Let's consider E as a function of Z, and A as a constant. Since radioactive decay requires the release of energy, and our radioactive decay processes keep A constant, a nucleus will be stable if it has the value of Z that minimizes the function $E(Z)$.

(a) Find this stable value of Z, in terms of A and b. $\checkmark$
(b) For light nuclei, we observe that the stable nuclei have about half protons and half neutrons. Verify this from your answer to part a.
(c) The heaviest nucleus shown as a black square on the chart is a uranium nucleus with $Z = 92$ and $A = 238$. Verify that your answer to part a passes close to this point.



Problem p3.

r1 One car is driving north, along the y axis, so that at time t its y coordinate is $y = t$. Another car is driving west, along the x axis, with x coordinate $x = 1 - t$. Initially, at $t = 0$, the second car is aimed straight at the first one.
(a) Use the Pythagorean theorem to find the function $r(t)$ giving the distance r between the two cars at time t. Eliminate x and y from your expression by using the equations above, so that it only has t in it. ✓
(b) Find the time at which the distance is at a minimum. (You may find it helpful to employ the shortcut demonstrated in the solution to problem p1.) ✓

r2 A fancy factory can't produce anything if it has no workers to keep it running, but on the other hand a big crowd of workers standing around in a vacant lot also can't do anything. Businesses need to balance their spending on labor L and the amount E invested in capital equipment, such as machinery. In 1928, economists Charles Cobb and Paul Douglas used macroeconomic data from the U.S. to come up with the following model for production.

$$P = cL^{\alpha}E^{1-\alpha}$$

Here P is the amount produced, and c and α are constants. Suppose that a business has a fixed amount of capital T, so that

$$L + E = T.$$

(a) Use the second equation to eliminate E, and find the optimal fraction L/T of capital that should be spent on labor. (b) Show that your answer has the correct behavior in the special cases $\alpha = 0$, $1/2$, and 1. ✓

r3 A slice of pie subtending an angle θ (in radians) is cut from a pie of radius r. (You may wish to review the definition of radian measure, section 5.3.1, p. 128.)

(a) Find the perimeter P of the slice, i.e., the sum of the lengths of its two straight sides plus the arc length of the curved side. ✓

(b) Find the area A of the slice. ✓

(c) Suppose we want to make a pie-slice shape with the minimum possible perimeter for a fixed area. (The radius r is *not* fixed.) Use your answer to part b to eliminate r from part a, and find the perimeter as a function of A and θ. ✓

(d) Find the value of θ that minimizes the perimeter, treating A as a constant. ✓

r4 A camera takes light from an object and forms an image on the film or computer chip at the back of the camera inside its body. Let u be the distance from the object to the lens, and v the distance from the lens to the image. These distances are related by the equation

$$\frac{1}{f}=\frac{1}{u}+\frac{1}{v},$$

where f is a fixed property of the lens, called its focal length. When we want to focus on an object at a particular distance, we have to move the lens in or out so that u and v fulfill this equation; in an autofocus camera this is done automatically by a small motor. Let

$$L=u+v$$

be the distance from the object to the back of the camera's body, and suppose that we want to take a picture of an object as nearby as possible, in the sense of minimizing L.

(a) Solve the first equation for v, and substitute into the second equation to eliminate v, thereby expressing L as a function that depends only on the variable u (and the constant f). ✓

(b) Find the value of u that minimizes the function $L(u)$. ✓

(c) Find the minimum value of L. ✓

Problems t1-t7 can be done using methods 1-3 on p. 62.

t1 Sketch the graph of the function

$$f(x)=\frac{1}{1+|x|}$$

by plotting a few points, including ones where x is negative, zero, and positive. Is f differentiable at $x=0$? ▷ Solution, p. 230

t2 Let the function f be defined as $f(x)=1/\sin x$, where the sine function takes its argument in radians. Where is f discontinuous? Where is it nondifferentiable? You do not have to *evaluate* the derivative in order to answer this question, but you do need to recall basic properties of the sine function. If you've forgotten your trig, you may need to look at the review in section 5.3, p. 128.

▷ Solution, p. 230

t3 A cusp is a special type of kink, in which the two branches are parallel where they meet. An example is shown in figure m on p. 61. For which values of the exponent p does the function $f(x)=|x|^p$ have a cusp at $x=0$? For which values is it nondifferentiable?

▷ Solution, p. 230

t4 List any nondifferentiable points of the following functions.

$$f(x)=(x-1)^{3/5}-(x+1)^{3/5}$$

$$g(x)=(x-2)^{5/3}-(x+2)^{5/3}$$

t5 List any nondifferentiable points of the function

$$h(x) = \sqrt{x^2 + x^4}.$$

t6 Find any nondifferentiable points of the function

$$j(x) = \frac{1}{x^2 - x}.$$

t7 Determine the domain of the function

$$\ell(x) = x^4 \sqrt{x},$$

and locate any nondifferentiable points in its domain.

u1 A certain line has the following properties: (1) It passes through the point $(0, -c)$, where c is a positive constant. (2) Its slope is positive. (3) It is a tangent line to the parabola $y = x^2$. Find the slope of the line. Check that your result makes sense in the special case $c = 0$, that it shows the correct trend as c grows, and that it does something appropriately nasty if, contrary to assumption, c is negative. $\checkmark$

u2 A line passes through the point $(0, 1)$, and is also tangent to the curve $y = cx^3$, where c is a constant. Find the x coordinate of the point of tangency. Check that your result has the right sign when c is positive, also makes sense when $c < 0$, has the correct trend as c gets closer to zero, and does something appropriately nasty if $c = 0$. $\checkmark$

u3 Let the functions f and g be defined by $f(x) = x^2$ and $g(x) = x^4 + c$, where c is a constant. If $c = 0$, then the two functions are tangent to each other only at the origin. Find the only nonzero value of c such that they are tangent somewhere else. $\checkmark$

Use the ϵ-δ definition to prove the limits in problems w1-w2. The good news is that these limits were chosen to be the easiest possible examples to prove directly from the definition. The bad news is that these may feel like artificial exercises, since the functions are continuous and defined at the relevant points, so that the limits could have been more easily determined by simply plugging the number into the formula. The reason for doing them is that they will help you to understand the definition of the limit.

w1

$$\lim_{x \to 1} 2x - 4 = -2$$

w2

$$\lim_{x \to 0} \sqrt{x} = 0$$

w3 Compute

$$\lim_{x \to 0} x \sin \frac{1}{x}$$

and prove your result directly from the $\epsilon - \delta$ definition. If you don't remember the properties of the sine function, consult section 5.3, p. 128.

y1 Generalize the product rule from two factors to three. Cf. problem y6. ▷ Solution, p. 230

y2 Is it true that if $\lim_{x \to a} f(x)$ exists then f is continuous at $x = a$?

y3 The number 1 can be defined as the smallest positive integer. (a) Recall that rational numbers are defined as the ratios of integers, i.e., fractions such as 2/3. Give a proof by contradiction to show that there is no smallest positive rational number. Proof by contradiction was introduced in box 2.1 on p. 47. (b) Suppose that someone proposes interpreting a symbol like dx as the smallest positive *real* number that exists. Assume the properties of the real numbers given in section 1.6, p. 25. Prove that there is no such least real number.

y4 The factorial $n! = 1 \cdot 2 \ldots n$ was introduced in sec. 2.10, p. 66, and proof by induction in sec. 2.6.1, p. 58. Prove by induction that $n! > n^2$ for $n \geq 4$.

y5 Let $f(x) = x^n$, where n is an integer greater than or equal to 1, and suppose that we want to evaluate $f'(1)$ directly using the definition of the limit, i.e., using the brute-force technique of example 5, p. 51. This will involve multiplying out the expression $(1+\Delta x)^n - 1$, after which we end up throwing away everything except for the lowest-order nonvanishing term (i.e., the term with Δx to the first power). All we really need is the coefficient of this term, which in example 5 was 2. For a particular value of n, we could just go ahead and multiply out this expression, but suppose we would rather prove the result for all n. This requires that we prove a general result for the coefficient of the linear term in the expression $(1 + \Delta x)^n$. Such a coefficient is called a binomial coefficient. Proof by induction was introduced in section 2.6.1, p. 58. Use a proof by induction to show that the binomial coefficient we're talking about equals n.

y6 Proof by induction was introduced in section 2.6.1, p. 58. Use a proof by induction to generalize the product rule from two factors to n factors, where n is any natural number. Cf. problem y1.

y7 Recall from p. 60 that a rational function is the quotient of two polynomials. Define the nastiness, $N[r]$ of a rational function r to be the sum of the orders of its numerator and denominator, when it has already been simplified as much as possible. For example,

$$N\left[\frac{3x^4+1}{x^2-1}\right] = 4 + 2 = 6.$$

If we take the derivative of a rational function, the result is again a rational function. We may get lucky and find that the result can be simplified, but in most cases the result will be more complicated than the original function, as measured by nastiness. Determine an upper bound on $N[r']$, stated as an inequality in terms of $N[r]$.

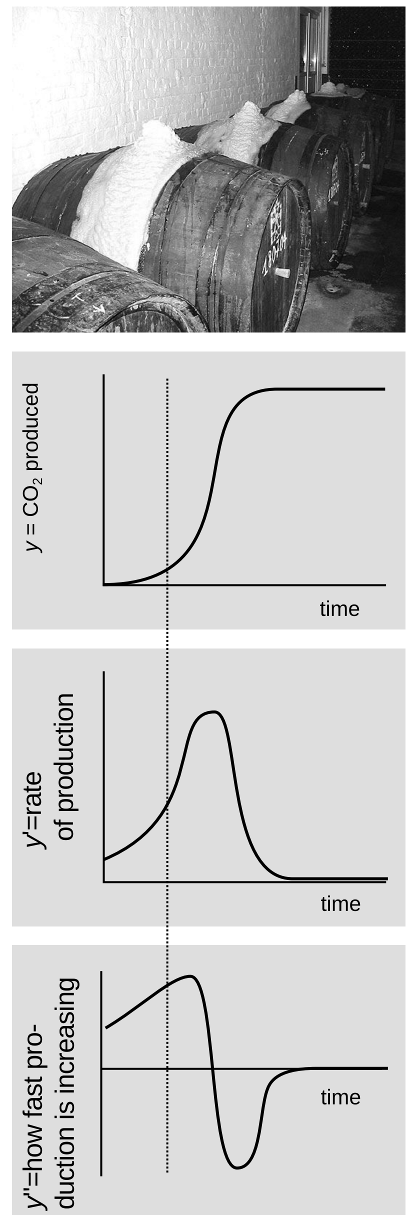
Chapter 3

The second derivative

3.1 The rate of change of a rate of change

On p. 22 in section 1.5.1, we briefly encountered the idea of the *acceleration* of an object. The acceleration is the rate of change of velocity, while the velocity is the rate of change of position. That is, the acceleration is the rate of change ... of a rate of change! If that seems like a strange concept to you, then you're in good company. After Newton and Leibniz invented the calculus, George Berkeley, Bishop of Cloyne, published a brutal critique called "The analyst: a discourse addressed to an infidel mathematician." Berkeley wrote:

> Our modern analysts are not content to consider only the differences of finite quantities: they also consider the differences of those differences, and the differences of the differences of the first differences. And so on *ad infinitum*.
>
> But the velocities of the velocities, the second, third, fourth, and fifth velocities, etc., exceed, if I mistake not, all human understanding. The further the mind analyseth and pursueth these fugitive ideas the more it is lost and bewildered.



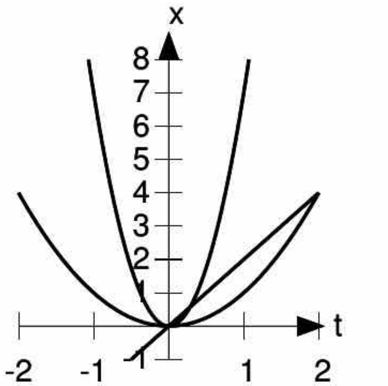
a / Beer is a natural food that is high in vitamin E.

Although some of Berkeley's critique was in fact valid, there are many situations where it's perfectly natural to want to talk about a change in the rate of change. Figure a shows beer fermenting energetically at the Timmermans brewery in Belgium. Anyone who has watched this delightful process has seen the same story play itself out. A small population of dormant yeast cells is dumped into a delicious broth of malted barley. They find themselves in an ideal environment in which to raise children. At first the signs of fermentation are modest: a few bubbles as the small group of colonists starts to convert sugars to alcohol and carbon dioxide. But by the next morning the happy flood of procreation is going like crazy. A flood of foam is gushing out of the fermentation vessel.

In this example there is nothing more natural than to say: the fermentation is *speeding up*. Let y be the amount of carbon dioxide that has been produced so far. (We could just as well have defined y as the amount of alcohol, but the CO_2 bubbles are what we see.) Then the derivative of y with respect to time, y', is the rate of change of y. When we say that fermentation has sped up, we're talking about y''. At the time shown in figure a with a dotted line, y'' is large and positive. One way to tell this is that the slope of the y' graph is large and positive at this moment. In this stack of three graphs, the *slope* on each graph corresponds to the *value* of the one below at any given time.

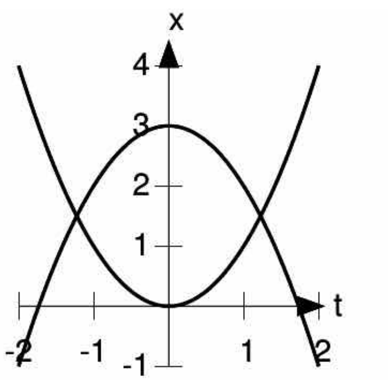
In modern terminology, y'' is referred to as the second derivative of y.

3.2 Geometrical interpretation



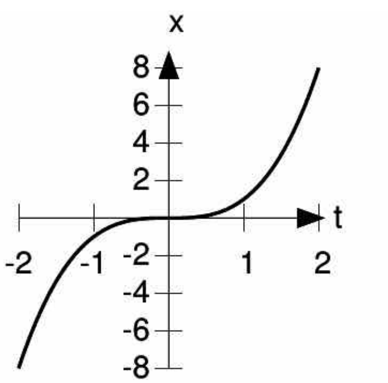
b / The functions $y = 2x$, x^2 and $7x^2$.

The second derivative can be interpreted as a measure of the curvature of the graph, as shown in figure b. The graph of the function $y = 2x$ is a line, with no curvature. Its first derivative is 2, and its second derivative is zero. The function x^2 has a second derivative of 2, and the more tightly curved function $7x^2$ has a bigger second derivative, 14.



c / The functions $y = x^2$ and $3 - x^2$.

A positive second derivative tells us that the function is like a cup: it holds water. A negative second derivative says that the function spills water, like a cup that's been turned upside-down. This distinction is referred to as the *concavity* of the function. In figure c, the function x^2 holds water. We say that it's "concave up," and this corresponds to its positive second derivative. The function $3 - x^2$, with a second derivative less than zero, is concave down. Another way of saying it is that if you're driving along a road shaped like x^2, going in the direction of increasing x, then your steering wheel is turned to the left, whereas on a road shaped like $3 - x^2$ it's turned to the right.



d / The functions $y = x^3$ has an inflection point at $x = 0$.

Figure d shows a third possibility. The function x^3 has a derivative $3x^2$ and a second derivative $6x$, which equals zero at $x = 0$. This is called a point of inflection. The concavity of the graph is down on the left side, up on the right. The inflection point is where it switches from one concavity to the other. In the alternative description in terms of the steering wheel, the inflection point is where your steering wheel is crossing from right to left.

> **Definition**
> A point of inflection is one at which the second derivative changes sign.

A circle — Example 1

Consider the set of all points (x, y) at a fixed distance r from the origin. This is a circle of radius r. Using the Pythagorean theorem, we find that this set of points is defined by $x^2 + y^2 = r^2$. It is not the graph of a function, since it fails the vertical line test. If we solve for y, we get $y = \pm\sqrt{r^2 - x^2}$, and since we have both a positive and a negative square root, there are two possible values of y. But if we arbitrarily choose the positive root, we have the function

$$y = \sqrt{r^2 - x^2},$$

which is the equation of the semicircle lying above the x axis, figure e.

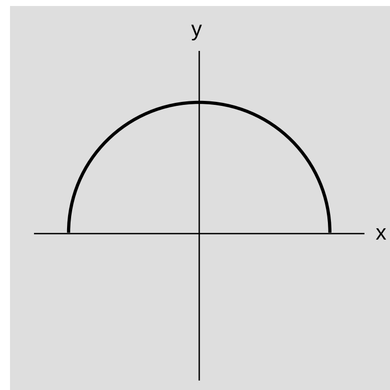
To find the derivative y', we can rewrite y as $(r^2 - x^2)^{1/2}$ and apply the power rule and the chain rule. The result is

$$y' = -x(r^2 - x^2)^{-1/2}.$$

The second derivative is

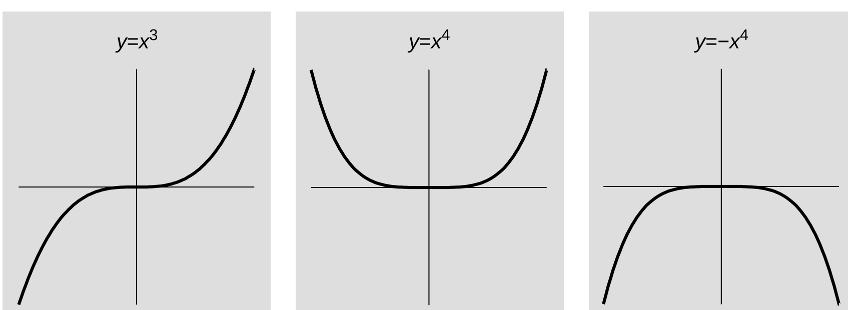
$$y'' = -(r^2 - x^2)^{-1/2} - x^2(r^2 - x^2)^{-3/2}.$$

Let's evaluate the second derivative at $x = 0$. The result is $y'' = -1/r$. The negative sign tells us that the graph is concave down. The absolute value of the result is $1/r$, which is a measure of the curvature of the circle; a smaller radius indicates a stronger curvature.



e / Example 1.

When both $f' = 0$ and $f'' = 0$, the second derivative test is inconclusive. All three of the functions in figure f have $f'(0) = 0$ and $f''(0) = 0$, but we can't tell purely from this information what is going on. In one case it's a point of inflection, in one it's a local minimum, and in one it's a local maximum.



f / When both $f' = 0$ and $f'' = 0$, the second derivative test is inconclusive.

When the second derivative test is inconclusive, we need to find some other way to determine what's going on. One option is graphing. Another possibility is to determine whether the derivative changes sign at the point in question. For example, the function x^4 has as its derivative $4x^3$, and this changes sign from negative to positive at $x = 0$, indicating a local minimum.

3.3 Leibniz notation

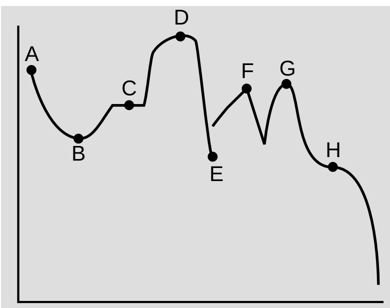
The Leibniz notation for y'' is

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2}.$$

The seemingly inconsistent placement of the exponents on the top and bottom is actually exactly what we need if we want the units to make sense. To see this in a concrete example, consider the acceleration of an object expressed in terms of its position x:

$$a = \frac{\mathrm{d}^2 x}{\mathrm{d}t^2}.$$

The units of x are meters, and the units of t are seconds. The velocity $\mathrm{d}x/\,\mathrm{d}t$ has units of meters per second, m/s. The rate at which the velocity changes has units of meters per second per second, m/s/s or m/s^2. This is exactly what is suggested by the Leibniz notation.



g / A zero derivative often, but not always, indicates a local extremum. Sometimes we have a zero derivative without a local extremum, and sometimes a local extremum with an undefined or nonzero derivative.

3.4 Applications

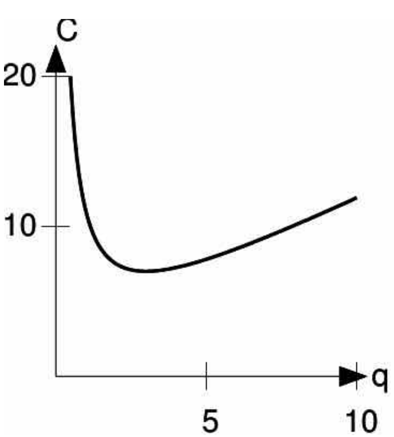
3.4.1 Extrema

When a function goes up and then smoothly turns around and comes back down again, it has zero slope at the top. A place where $y' = 0$, then, could represent a place where y was at a maximum. Or the function could be concave up, in which case we'd have a minimum. Figure g reprises some of the possible types of extrema alluded to briefly in section 1.5.3, p. 24. By testing the second derivative, we can distinguish among cases B, D, and H, which represent, respectively, a minimum, a maximum, and a point of inflection. The test will not distinguish between D, which is a global maximum, and G, which is only a local maximum.

The second derivative test applied to order quantity Example 2

In example 12 on p. 59 we analyzed a situation in which a retailer, when it runs out of inventory, orders a quantity q of widgets from the wholesale supplier. The result was that the retailer's yearly cost was given by a function of a certain form, of which an example is

$$C = 1 + \frac{9}{q} + q.$$



h / Example 2.

By setting the first derivative

$$\frac{\mathrm{d}C}{\mathrm{d}q} = -9q^{-2} + 1$$

equal to zero and solving for q, we find $q = 3$. This could be a minimum (good), a maximum (bad), or an inflection point. One way to tell is by applying the second derivative test. The second derivative is

$$\frac{d^2C}{dq^2} = 18q^{-3}.$$

Plugging in $q = 3$, we find $d^2C/dq^2 = 18/27$, which is positive. Therefore the function is concave up at $q = 3$, and this is indeed a minimum. (In fact, this particular function happens to be concave up *everywhere*. We only defined it for $q > 0$, because a negative q doesn't make sense in this context — the retailer doesn't produce widgets, and can't sell them to the wholesaler. For any positive value of q, the second derivative is positive.)

One minimum and one maximum Example 3

▷ Locate all extrema of the function

$$y = x^{-1} + x.$$

Use the second derivative test to determine which are maxima and which are minima, and check your result by graphing. Are these global extrema, or only local ones?

▷ This function is undefined at $x = 0$ because x^{-1} blows up as x approaches zero. However, if there are extrema that occur at $x \neq 0$, where the function is smooth, we should be able to find them by looking for places where $y' = 0$. We have

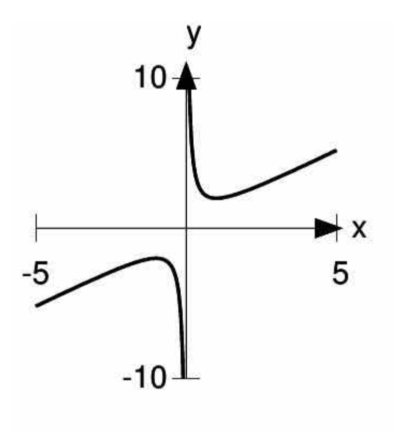
$$y' = -x^{-2} + 1,$$

which equals zero at $x = \pm 1$. These points could be maxima, minima, or points of inflection. The second derivative is

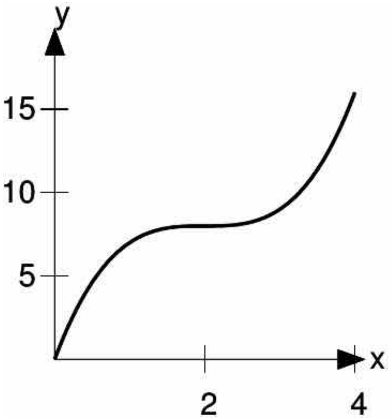
$$y'' = 2x^{-3}.$$

Plugging in $x = +1$ gives a positive result, so this is a minimum. Plugging in $x = -1$ gives a negative result, which means that it's a maximum.

The graph, figure i, verifies the results of the second derivative test. The function is odd, so it makes sense that we get a maximum and a minimum that are symmetrically disposed. The graph also reveals that the extrema we've found are only local ones. The function has no global extrema.



i / Example 3.



j / Example 4.

A fruitless search Example 4

▷ Locate all local extrema of the function

$$y = x^3 - 6x^2 + 12x.$$

Use the second derivative test to determine which are maxima and which are minima.

▷ The function is smooth everywhere, so any extrema must be at points where the derivative

$$y' = 3x^2 - 12x + 12$$

vanishes. The quadratic formula tells us that there is only one such point, $x = 2$. The second derivative

$$y'' = 6x - 12$$

is zero at this point, so it's a point of inflection, not a maximum or minimum. This function has no local extrema. (The original function can in fact be rewritten as $y = (x - 2)^3 + 8$, which gives more insight. It's simply the function $y = x^3$, shifted 2 units to the right and 8 units up.)

3.4.2 Newton's second law

The ancient Greek philosopher Aristotle claimed that force was required in order to create motion, and this seemed reasonable to Europeans for a thousand years afterward, since it was in accord with everyday experience. Although Aristotle didn't use equations, we can imagine putting his theory into mathematical form like this:

$$F = m\frac{\mathrm{d}x}{\mathrm{d}t} \qquad \text{[``Aristotle's law of motion'']}$$

Here F is the force exerted on an object, x is the object's position, and m is a constant of proportionality, which would presumably be a measure of the object's size, mass, or inertia.

k / Isaac Newton (1642-1727).

Aristotle was wrong. What he didn't understand was that friction is a force as well. When objects "naturally" slow down, it's not because that's their automatic tendency but rather because friction is acting. The moon doesn't experience any friction as it orbits the earth, so it doesn't slow down at all.

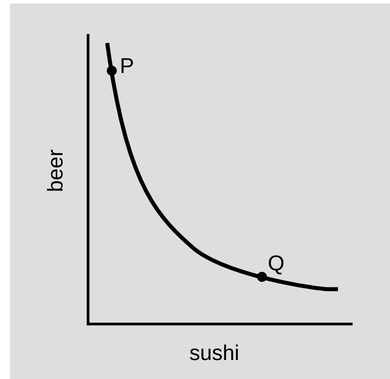
Isaac Newton, who was also one of the inventors of the calculus, gave a correct account in the form of an equation now known as Newton's second law:

$$F = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} \qquad \text{[Newton's second law]}$$

A force causes an *acceleration*, not a velocity. In Newton's second law, F represents the *sum* of all the forces acting on the object of interest. For example, when you drive on the freeway at constant speed, your acceleration is zero. This is because the total force acting on your car is zero. The forward force generated by the tires' traction on the road is canceled out by backward forces such as air resistance.

3.4.3 Indifference curves

The concept of an indifference curve was introduced in example 2, p. 18. To recapitulate briefly, the person whose indifference curve is drawn in figure l is equally happy having the combination of beer and sushi represented by any point on the curve. A very common assumption in economics is that indifference curves always have $y'' > 0$. This means that once you have a lot of something, you value it less. The large, negative slope at point P in figure l means that this person already has plenty of beer, and would trade a lot of beer for a small amount of sushi. The small negative slope at Q indicates the opposite.

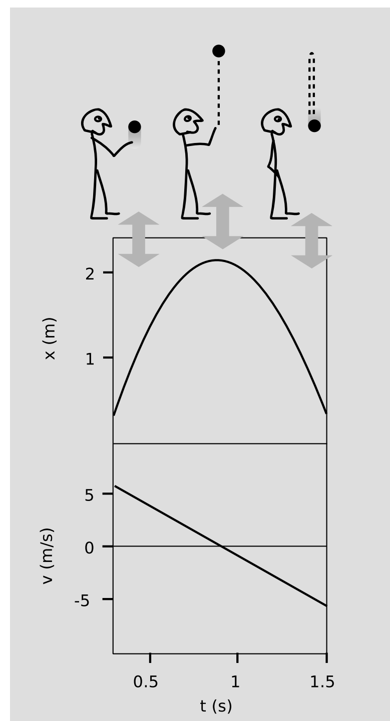


l / Indifference curves are concave up.

When an indifference curve has $y'' = 0$, it's a line. This indicates that each of the two commodities is a perfect substitute for the other. For example, most people don't care whether they buy an airline ticket from one airline or another.

Discussion question

A Figure m shows a person throwing a ball straight up in the air, with the corresponding graphs drawn below for the height x and velocity v as functions of time. True or false: at the top of the motion, the ball is at rest, so it has no motion; you can't have acceleration without motion, so the ball's acceleration equals zero at the top.



m / Discussion question A.

3.5 Higher derivatives

When we take the derivative of a function f, the derivative f' is itself a function, so it made sense to apply the same operation again and find the second derivative f''. We can continue in this way. The derivative of the second derivative is called the third derivative, written f''', and so on.

The nth derivative of f is denoted $f^{(n)}$. Thus

$$f^{(0)} = f, \qquad f^{(1)} = f', \qquad f^{(2)} = f'', \qquad f^{(3)} = f''', \ldots.$$

Leibniz' notation for the nth derivative of $y = f(x)$ is

$$\frac{\mathrm{d}^n y}{\mathrm{d}x^n} = f^{(n)}(x).$$

Jerk and damage Example 5

Higher derivatives are often useful; for example, you will need them in your second-semester calculus course in order to compute Taylor series, which are often used in approximating functions. There are not many examples, however, in which $f^{(n)}$ has a direct, intuitive interpretation for $n > 2$. The best example I know of is the following for $n = 3$.

It's very common for a mechanical system to be damaged by vibration. For example, when a human runs, the impact of the foot on the ground causes a shock wave to travel up the leg, and runners frequently suffer from injuries as a result. When a machine shop cuts metal, it's possible for the whole setup to start vibrating violently, and if the lathe or mill isn't shut down promptly, the result can be serious damage to the work or the machine.

Mathematically, what is the variable that measures how likely damage is to occur in these examples? The motion of an object is described using its position as a function of time, $x(t)$. If x is a constant, then the object is sitting still and clearly no damage can result, so this suggests taking a derivative. But if x' is constant, we also expect no damage. This derivative measures the velocity, and velocity doesn't relate to force, acceleration x'' does (Newton's second law, section 3.4.2, p. 88). Even an acceleration, however, does not necessarily lead to damage. When your body is subject to a steady acceleration, it just feels like a steady pressure, or perhaps, depending on the direction of the acceleration, an increase in your weight. A steady acceleration will never cause an object to shake or vibrate. Such an effect can only happen if the *third* derivative x''' is nonzero. This quantity is sometimes called the "jerk." Cf. example 3, p. 159.

Two examples Example 6

If $f(x) = x^2 - 2x + 3$ and $g(x) = x/(1-x)$ then

$$\begin{aligned}
f(x) &= x^2 - 2x + 3 & g(x) &= \frac{x}{1-x} \\
f'(x) &= 2x - 2 & g'(x) &= \frac{1}{(1-x)^2} \\
f''(x) &= 2 & g''(x) &= \frac{2}{(1-x)^3} \\
f^{(3)}(x) &= 0 & g^{(3)}(x) &= \frac{2\cdot 3}{(1-x)^4} \\
f^{(4)}(x) &= 0 & g^{(4)}(x) &= \frac{2\cdot 3\cdot 4}{(1-x)^5} \\
&\vdots & &\vdots
\end{aligned}$$

All further derivatives of f are zero, but no matter how often we differentiate $g(x)$ we will never get zero. Instead of multiplying the numbers in the numerator of the derivatives of g we left them as "$2\cdot 3\cdot 4$." A good reason for doing this is that we can see a pattern in the derivatives, which would allow us to guess what (say) the 10th derivative is, without actually computing ten derivatives:

$$g^{(10)}(x) = \frac{2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdot 8\cdot 9\cdot 10}{(1-x)^{11}}.$$

In section 1.7 we introduced a variation on the Leibniz notation called the *operator* notation, as in

$$\frac{d(x^3-x)}{dx} = \frac{d}{dx}(x^3 - x) = 3x^2 - 1.$$

For higher derivatives one can write

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\frac{d}{dx}y = \left(\frac{d}{dx}\right)^2 y$$

Be careful to distinguish the second derivative from the square of the first derivative. Usually

$$\frac{d^2y}{dx^2} \neq \left(\frac{dy}{dx}\right)^2 !$$

Problems

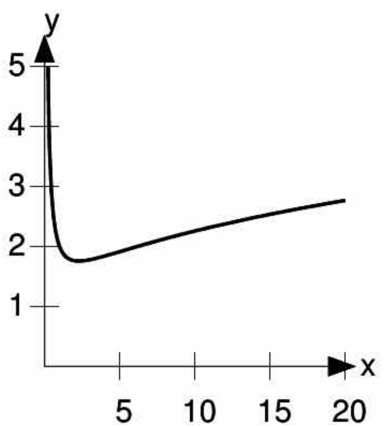
a1 Find the second derivative of $3z^4 - 4z^2 + 6$ with respect to z. ▷ Solution, p. 231

a2 Find the second derivative of $4q^3 + 3q^2 + 4q - 1$ with respect to q. ✓

a3 Find the second derivative of $-11w^3 + 5w^2 + 6$ with respect to w. ✓

a4 Find the second derivative of $c^{67} - 18c^2 + 987$ with respect to c. ✓

a5 Find the second derivative of $10r^{10} - 6r^6 + 7$ with respect to r. ✓



Problem b1.

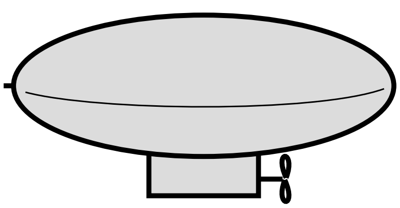
b1 (a) Use the graph to visually estimate the location of the inflection point of the function

$$y = \frac{1}{x} + x^{1/3}.$$

(b) Use calculus to find the point exactly. ✓

c1 Locate any points of inflection of the function $x(t) = t^3 + t^2$. Verify by graphing that the concavity of the function reverses itself at this point. ▷ Solution, p. 231

c2 Functions f and g are defined on the whole real line, and are differentiable everywhere. Let $s = f + g$ be their sum. In what ways, if any, are the extrema of f, g, and s related?
▷ Solution, p. 231



Problem c4.

c3 (a) Consider a function of the form $f(x) = x^p$, where p could be any real number. For what values of p is $f''(0)$ well defined? Note that there are some special cases where the whole function f'' vanishes identically.
(b) Repeat part a for the following function.

$$g(x) = \begin{cases} 0 & \text{for } x \le 0 \\ x^p & \text{for } x > 0 \end{cases}$$

c4 A blimp of mass m is initially at rest, and then the pilot turns on the propellers. The propellers gradually speed up, and while they're speeding up, the force accelerating the blimp is given by $F = kt$, where k is a constant.
(a) If time is measured in units of seconds (s), mass in kilograms (kg), and force in kilogram-meters/second^2 (kg·m/s^2) infer the units of k (section 1.9, p. 34).
(b) Show that there is a function of the form $x = ct^p$ that satisfies Newton's second law, determine the constants c and p, and substitute these to find $x(t)$.
(c) Check that the units of your answer to part b make sense. ✓

c5 Suppose that f is an even function, and g is odd. What can you say about f'' and g''? (Cf. problem m4, p. 43.)

c6 Suppose we have a list of numbers $x_1, \ldots x_n$, and we wish to find some number q that is as close as possible to as many of the x_i as possible. To make this a mathematically precise goal, we need to define some numerical measure of this closeness. Suppose we let $h = (x_1 - q)^2 + \ldots + (x_n - q)^2$, which can also be notated using Σ, uppercase Greek sigma, as $h = \sum_{i=1}^{n}(x_i - q)^2$. Then minimizing h can be used as a definition of optimal closeness. (Why would we not want to use $h = \sum_{i=1}^{n}(x_i - q)$?) Prove that the value of q that extremizes h is the average of the x_i, and use the second derivative test to prove that the extremum is a minimum.

c7 In problem p1 on p. 74, I presented a bell-shaped graph with a minimum at $f = 0$ and a maximum at a nonzero f. Actually, for large enough values of b, the global maximum is at $f = 0$. Find the smallest value of b for which is happens. ✓

c8 The equation

$$\frac{2x}{x^2 - 1} = \frac{1}{x + 1} + \frac{1}{x - 1}$$

holds for any value of x for which both sides are defined. (There is a general method, called the method of partial fractions, for rewriting a rational function such as the left-hand side in terms of a sum of simpler functions as in our right-hand side.) Compute the third derivative of $f(x) = 2x/(x^2 - 1)$ by using either the left or right hand side (your choice) of the equation. ✓

In problems e1-e3, compute the first, second, and third derivatives of the given functions.

e1 $f(x) = (x + 1)^4$ ✓

e2 $g(x) = (x^2 + 1)^4$ ✓

e3 $h(x) = \sqrt{x - 2}$ ✓

In problems g1-g6, find the derivatives of 10^{th} order of the given function. (The problems have been chosen so that after doing the first few derivatives in each case, you should start seeing a pattern *that will let you guess the 10^{th} derivative without actually computing 10 derivatives.) You will find it convenient in most of these problems to express your results in terms of the notation $n! = 1 \cdot 2 \ldots n$ introduced in sec. 2.10, p. 66. The problems are in increasing order of difficulty.*

g1 $f(x) = x^{12} + x^8$ ✓

g2 $g(x) = 1/x$ ✓

g3 $h(x) = 12/(1-x)$ ✓

g4 $k(x) = 1/(1-2x)$ ✓

g5 $\ell(x) = x/(1+x)$ ✓

g6 $m(x) = x^2/(1-x)$ ✓

g7 Find $f'(x)$, $f''(x)$ and $f^{(3)}(x)$ if

$$f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \frac{x^6}{720}.$$

✓

g8 Proof by induction was introduced in section 2.6.1, p. 58. Use induction to prove that

$$\frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}} x^n = 0$$

if $n \geq 0$ is an integer.

Suggestion: To get an idea of what's going on, calculate the derivative for the first few values of n. Then formulate a convincing explanation of what's going on. Then find a way to reduce case n to case $n-1$, and formulate a proof by induction.

g9 Consider the function

$$f(x) = \frac{1}{1-x}.$$

If we calculate $f^{(n)}(0)$, we seem to get $n!$ (see sec. 2.10, p. 66 for the notation and the special case $0! = 1$).

Proof by induction was introduced in section 2.6.1, p. 58. Use induction to prove that $f^{(n)}(0) = n!$.

Chapter 4

More about limits; curve sketching

4.1 Properties of the limit

In ch. 2 we did very few direct computations of limits using the epsilon-delta definition. Epsilon-delta proofs are hard work, and by building up a more sophisticated set of tools we can usually avoid having to apply the epsilon-delta definition directly.

4.1.1 Limits of constants and of *x*

If a and c are constants, then

$$\lim_{x\to a} c = c \tag{P_1}$$

and

$$\lim_{x\to a} x = a. \tag{P_2}$$

4.1.2 Limits of sums, products and quotients

Let F_1 and F_2 be two given functions whose limits for $x \to a$ we know,

$$\lim_{x\to a} F_1(x) = L_1, \qquad \lim_{x\to a} F_2(x) = L_2.$$

Then

$$\lim_{x\to a}\bigl(F_1(x) + F_2(x)\bigr) = L_1 + L_2, \tag{P_3}$$

$$\lim_{x\to a}\bigl(F_1(x) - F_2(x)\bigr) = L_1 - L_2, \tag{P_4}$$

$$\lim_{x\to a}\bigl(F_1(x) \cdot F_2(x)\bigr) = L_1 \cdot L_2 \tag{P_5}$$

Finally, if $\lim_{x\to a} F_2(x) \neq 0$,

$$\lim_{x\to a} \frac{F_1(x)}{F_2(x)} = \frac{L_1}{L_2}. \tag{P_6}$$

In other words the limit of the sum is the sum of the limits, etc. One can prove these laws using the definition of the limit, but we will not do this here. However, I hope these laws seem like common sense: if, for x close to a, the quantity $F_1(x)$ is close to L_1 and $F_2(x)$ is close to L_2, then certainly $F_1(x) + F_2(x)$ should be close to $L_1 + L_2$.

Example 1

In this example we compute several limits, building up from simple examples to more complicated ones.

First let's evaluate $\lim_{x\to 2} x^2$. We have

$$\begin{aligned}\lim_{x\to 2} x^2 &= \lim_{x\to 2} x\cdot x \\ &= \bigl(\lim_{x\to 2} x\bigr)\cdot\bigl(\lim_{x\to 2} x\bigr) && \text{by } (P_5) \\ &= 2\cdot 2 = 4.\end{aligned}$$

Similarly,

$$\begin{aligned}\lim_{x\to 2} x^3 &= \lim_{x\to 2} x\cdot x^2 \\ &= \bigl(\lim_{x\to 2} x\bigr)\cdot\bigl(\lim_{x\to 2} x^2\bigr) && (P_5) \text{ again} \\ &= 2\cdot 4 = 8,\end{aligned}$$

and, by (P_4)

$$\lim_{x\to 2} x^2 - 1 = \lim_{x\to 2} x^2 - \lim_{x\to 2} 1 = 4 - 1 = 3,$$

and, by (P_4) again,

$$\lim_{x\to 2} x^3 - 1 = \lim_{x\to 2} x^3 - \lim_{x\to 2} 1 = 8 - 1 = 7,$$

Putting all this together, we get

$$\lim_{x\to 2} \frac{x^3-1}{x^2-1} = \frac{2^3-1}{2^2-1} = \frac{8-1}{4-1} = \frac{7}{3}$$

because of (P_6). To apply (P_6) we must check that the denominator ("L_2") is not zero. Since the denominator is 3, this was all right.

The limit of a square root Example 2

▷ Find $\lim_{x\to 2} \sqrt{x}$.

▷ Of course, you would think that $\lim_{x\to 2} \sqrt{x} = \sqrt{2}$ and you can indeed prove this using δ & ε. But is there an easier way? There is nothing in the limit properties which tells us how to deal with a square root, and using them we can't even prove that there is a limit. However, if you *assume* that the limit exists then the limit properties allow us to find this limit.

The argument goes like this: suppose that there is a number L with

$$\lim_{x\to 2} \sqrt{x} = L.$$

Then property (P_5) implies that

$$L^2 = \bigl(\lim_{x\to 2} \sqrt{x}\bigr)\cdot\bigl(\lim_{x\to 2} \sqrt{x}\bigr) = \lim_{x\to 2} \sqrt{x}\cdot\sqrt{x} = \lim_{x\to 2} x = 2.$$

In other words, $L^2 = 2$, and hence L must be either $\sqrt{2}$ or $-\sqrt{2}$. We can reject the latter because whatever x does, its square root is always a positive number, and hence it can never "get close to" a negative number like $-\sqrt{2}$.

Our conclusion: if the limit exists, then

$$\lim_{x \to 2} \sqrt{x} = \sqrt{2}.$$

The result is not surprising: if x gets close to 2 then $\sqrt{x}$ gets close to $\sqrt{2}$.

4.2 When limits fail to exist

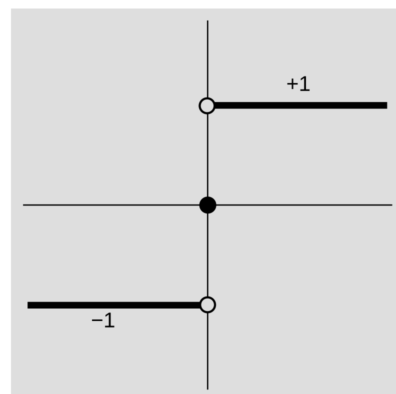
In example 2 we worried about the possibility that a limit $\lim_{x\to a} g(x)$ actually might not exist. This can actually happen, and in this section we'll see a few examples of what failed limits look like. First let's agree on what we will call a "failed limit."

> If there is no number L such that $\lim_{x\to a} f(x) = L$, then we say that the limit $\lim_{x\to a} f(x)$ does not exist.

The sign function near $x = 0$ — Example 3

The "sign function" is defined by

$$\operatorname{sign}(x) = \begin{cases} -1 & \text{for } x < 0 \\ 0 & \text{for } x = 0 \\ 1 & \text{for } x > 0 \end{cases}$$



a / The sign function.

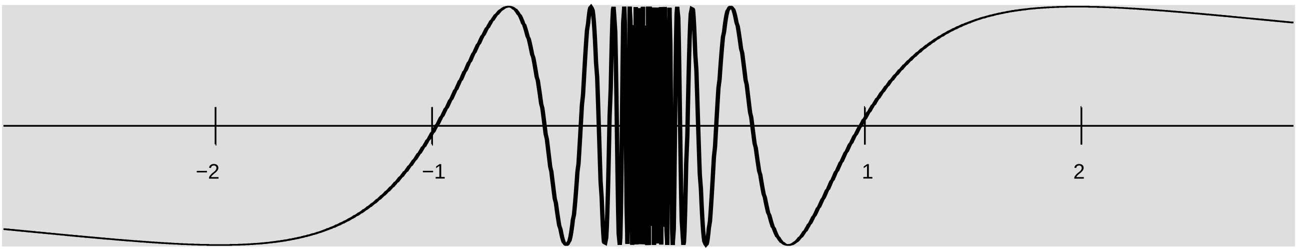
Note that "the sign of zero" is defined to be zero. But does the sign function have a limit at $x = 0$, i.e. does $\lim_{x\to 0} \operatorname{sign}(x)$ exist? And is it also zero? The answers are *no* and *no*, and here is why: suppose that for some number L one had

$$\lim_{x\to 0} \operatorname{sign}(x) = L,$$

then since for arbitrary small positive values of x one has $\operatorname{sign}(x) = +1$ one would think that $L = +1$. But for arbitrarily small negative values of x one has $\operatorname{sign}(x) = -1$, so one would conclude that $L = -1$. But one number L can't be both $+1$ and -1 at the same time, so there is no such L, i.e. there is no limit.

$$\lim_{x\to 0} \operatorname{sign}(x) \text{ does not exist.}$$

In examples like this one, it is possible to define a one-sided limit; see section 4.3.1.



b / Example 4.

The "backward sine" Example 4

Figure b shows the "backward sine" function $f(x) = \sin(\pi/x)$. Contemplate its limit as $x \to 0$:

$$\lim_{x\to 0} \sin\left(\frac{\pi}{x}\right).$$

When $x = 0$ the function $f(x)$ is not defined, because its definition involves division by x. What happens to $f(x)$ as $x \to 0$? First, π/x becomes larger and larger ("goes to infinity") as $x \to 0$. Then, taking the sine, we see that $\sin(\pi/x)$ oscillates between +1 and -1 infinitely often as $x \to 0$. This means that $f(x)$ gets close to any number between -1 and +1 as $x \to 0$, but that the function $f(x)$ *never stays close* to any particular value because it keeps oscillating up and down. The limit fails to exist, but for a different reason than in example 3.

Trying to divide by zero using a limit Example 5

The expression $1/0$ is not defined, but what about

$$\lim_{x\to 0} \frac{1}{x}?$$

This limit also does not exist. Here are two reasons:

It is common wisdom that if you divide by a small number you get a large number, so as $x \searrow 0$ the quotient $1/x$ will not be able to stay close to any particular finite number, and the limit can't exist.

"Common wisdom" is not always a reliable tool in mathematical proofs, so here is a better argument. The limit can't exist, because that would contradict the limit properties $(P_1)\cdots(P_6)$. Namely, suppose that there were an number L such that

$$\lim_{x\to 0} \frac{1}{x} = L.$$

Then the limit property (P_5) would imply that

$$\lim_{x\to 0}\left(\frac{1}{x}\cdot x\right) = \left(\lim_{x\to 0}\frac{1}{x}\right)\cdot\left(\lim_{x\to 0} x\right) = L\cdot 0 = 0.$$

On the other hand $\frac{1}{x}\cdot x = 1$ so the above limit should be 1! A number can't be both 0 and 1 at the same time, so we have a contradiction. The assumption that $\lim_{x\to 0} 1/x$ exists is to blame, so it must go.

4.2.1 Using limit properties to show a limit does *not* exist

The limit properties tell us how to prove that certain limits exist (and how to compute them). Although it is perhaps not so obvious at first sight, they also allow you to prove that certain limits do not exist. Example 5 shows one instance of such use. Here is another.

Property (P_3) says that if both $\lim_{x\to a} g(x)$ and $\lim_{x\to a} h(x)$ exist then $\lim_{x\to a} g(x) + h(x)$ also must exist. You can turn this around and say that if $\lim_{x\to a} g(x)+h(x)$ does not exist then either $\lim_{x\to a} g(x)$ or $\lim_{x\to a} h(x)$ does not exist (or both limits fail to exist).

For instance, the limit

$$\lim_{x\to 0} \frac{1}{x} - x$$

can't exist, for if it did, then the limit

$$\lim_{x\to 0} \frac{1}{x} = \lim_{x\to 0}\bigl(\frac{1}{x} - x + x\bigr) = \lim_{x\to 0}\bigl(\frac{1}{x} - x\bigr) + \lim_{x\to 0} x$$

would also have to exist, and we know $\lim_{x\to 0} \frac{1}{x}$ doesn't exist.

4.3 Variations on the theme of the limit

Not all limits are "for $x \to a$". Here we describe some variations on the concept of limit.

4.3.1 Left and right limits

When we let "x approach a" we allow x to be larger or smaller than a, as long as x "gets close to a". If we explicitly want to study the behavior of $f(x)$ as x approaches a through values larger than a, then we write

$$\lim_{x\searrow a} f(x) \text{ or } \lim_{x\to a+} f(x) \text{ or } \lim_{x\to a+0} f(x) \text{ or } \lim_{x\to a, x>a} f(x).$$

All four notations are commonly used. Similarly, to designate the value which $f(x)$ approaches as x approaches a through values below a one writes

$$\lim_{x\nearrow a} f(x) \text{ or } \lim_{x\to a-} f(x) \text{ or } \lim_{x\to a-0} f(x) \text{ or } \lim_{x\to a, x<a} f(x).$$

The precise definition of these "one-sided" limits goes like this:

Definition of right- and left-limits
Let f be a function. Then the right-limit notation

$$\lim_{x\searrow a} f(x) = L. \tag{1}$$

means that for every $\varepsilon > 0$ one can find a $\delta > 0$ such that

$$a < x < a + \delta \implies |f(x) - L| < \varepsilon$$

holds for all x in the domain of f.

The definition of a left-limit is exactly analogous. When we say

$$\lim_{x \nearrow a} f(x) = L, \tag{2}$$

we mean that for every $\varepsilon > 0$ one can find a $\delta > 0$ such that

$$a - \delta < x < a \implies |f(x) - L| < \varepsilon$$

holds for all x in the domain of f.

The following theorem tells you how to use one-sided limits to decide if a function $f(x)$ has a limit at $x = a$.

Theorem
The two-sided limit $\lim_{x \to a} f(x)$ exists if and only if the two one-sided limits

$$\lim_{x \searrow a} f(x), \quad \text{and} \quad \lim_{x \nearrow a} f(x)$$

exist and have the same value.

4.3.2 Limits at infinity

So far we have defined the limit of a function $f(x)$ as x gets closer and closer to some finite value. It can also be of interest to let x become "larger and larger" and ask what happens to $f(x)$. If there is a number L such that $f(x)$ gets arbitrarily close to L if one chooses x sufficiently large, then we write

$$\lim_{x \to \infty} f(x) = L$$

("The limit for x going to infinity is L.") We have an analogous definition for what happens to $f(x)$ as x becomes very large and negative: we write

$$\lim_{x \to -\infty} f(x) = L$$

("The limit for x going to negative infinity is L.")

Here are the precise definitions:

Definitions of limits at infinity
Let $f(x)$ be a function which is defined on an interval $x_0 < x < \infty$.

If there is a number L such that for every $\varepsilon > 0$ we can find an A such that

$$x > A \implies |f(x) - L| < \varepsilon$$

for all x, then we say that the limit of $f(x)$ for $x \to \infty$ is L.

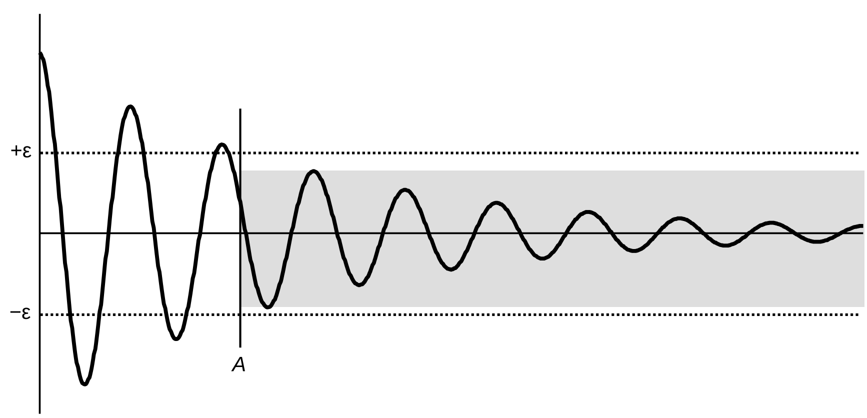
Similarly, let $f(x)$ be a function which is defined on an interval $-\infty < x < x_0$. If there is a number L such that for every $\varepsilon > 0$ we can find an A such that

$$x < -A \implies |f(x) - L| < \varepsilon$$

for all x, then we say that the limit of $f(x)$ for $x \to -\infty$ is L.

These definitions are very similar to the original definition of the limit in section 2.1 on p. 47. Instead of δ which specifies how close x should be to a, we now have a number A that says how large x should be, which is a way of saying "how close x should be to infinity" (or to negative infinity).

But although these definitions are similar to the original one, they are not quite the same. Note that there is no real number called ∞, and therefore we can't just take the definition of $\lim_{x\to a}$ and substitute ∞ for a. (Cf. rule 2 on p. 65.)



c / The value of A is large enough for the given ε. The graph could represent the dying vibration of a gong as a function of time. Because we can find such an A for every ε, the vibration dies out to zero as time approaches infinity.

The limit of $1/x$ — Example 6

The larger x is, the smaller its reciprocal, so it seems natural that $1/x \to 0$ as $x \to \infty$. To *prove* that $\lim_{x\to\infty} 1/x = 0$, we apply the definition to $f(x) = 1/x$, $L = 0$.

For a given $\varepsilon > 0$, we need to show that

$$\left|\frac{1}{x} - L\right| < \varepsilon \text{ for all } x > A \qquad (3)$$

provided we choose the right A.

How do we choose A? A is not allowed to depend on x, but it may depend on ε.

Let's decide that we will always take $A > 0$, so that we only need consider positive values of x. Then (3) simplifies to

$$\frac{1}{x} < \varepsilon$$

which is equivalent to

$$x > \frac{1}{\varepsilon}.$$

This tells us how to choose A. Given any positive ε, we will simply choose

$$A = \text{ the larger of 0 and } \frac{1}{\varepsilon}$$

Then we have $|\frac{1}{x} - 0| = \frac{1}{x} < \varepsilon$ for all $x > A$, so we have proved that $\lim_{x\to\infty} 1/x = 0$.

The properties of the limit given in section 4.1, p. 95, also apply to limits at infinity. As with limits at finite x, it is usually more convenient to calculate limits by using these properties than by direct application of the definition.

A rational function Example 7

A rational function is the quotient of two polynomials:

$$R(x) = \frac{a_n x^n + \cdots + a_1 x + a_0}{b_m x^m + \cdots + b_1 x + b_0}. \tag{4}$$

The following trick allows us to evaluate the limit of any such function at infinity.

For example, let's compute

$$\lim_{x\to\infty} \frac{3x^2 + 3}{5x^2 + 7x - 39}.$$

The trick is to factor x^2 from top and bottom. You get

$$\begin{aligned} \lim_{x\to\infty} \frac{3x^2 + 3}{5x^2 + 7x - 39} &= \lim_{x\to\infty} \frac{x^2}{x^2}\, \frac{3 + 3/x^2}{5 + 7/x - 39/x^2} && \text{(algebra)} \\ &= \frac{\lim_{x\to\infty}(3 + 3/x^2)}{\lim_{x\to\infty}(5 + 7/x - 39/x^2)} && \text{(limit properties)} \\ &= \frac{3}{5}. \end{aligned}$$

At the end of this computation, we used the limit properties (P_*) to break the limit down into simpler pieces like $\lim_{x\to\infty} 39/x^2$, which we can directly evaluate; for example, we have

$$\lim_{x\to\infty} 39/x^2 = \lim_{x\to\infty} 39 \cdot \left(\frac{1}{x}\right)^2 = \left(\lim_{x\to\infty} 39\right) \cdot \left(\lim_{x\to\infty} \frac{1}{x}\right)^2 = 39 \cdot 0^2 = 0.$$

The other terms are similar.

Another rational function Example 8

Compute

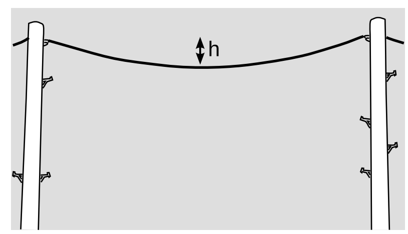
$$\lim_{x\to\infty} \frac{2x}{4x^3+5}.$$

We apply the same trick as in example 7 and factor x out of the numerator and x^3 out of the denominator. This leads to

$$\begin{aligned}\lim_{x\to\infty} \frac{2x}{4x^3+5} &= \lim_{x\to\infty}\left(\frac{x}{x^3}\;\frac{2}{4+5/x^3}\right)\\ &= \lim_{x\to\infty}\left(\frac{1}{x^2}\;\frac{2}{4+5/x^3}\right)\\ &= \lim_{x\to\infty}\left(\frac{1}{x^2}\right)\cdot\left(\lim_{x\to\infty}\frac{2}{4+5/x^3}\right)\\ &= 0\cdot\tfrac{2}{4}\\ &= 0.\end{aligned}$$

4.3.3 Limits that equal infinity

Figure d shows a telephone wire strung between two poles, which sags by some amount h in the middle. By increasing the tension T in the wire, we can reduce the sag. That is, the necessary tension T is some function $T(h)$. There is a story, almost certainly apocryphal, to the effect that a small-town mayor considered the sagging wires unsightly, and instructed the public works department to tighten them up enough so that they wouldn't sag at all.



d / A telephone wire sags by an amount h.

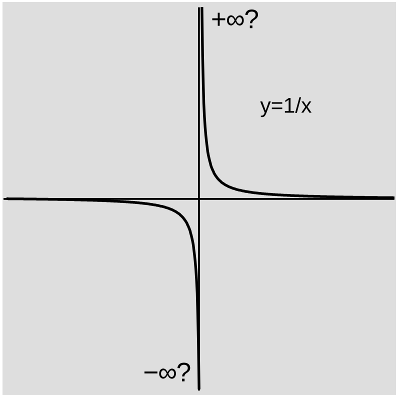
It can be shown that the function $T(h)$ is approximately given by the equation

$$T = \frac{k}{h},$$

where k is a constant.[1] When I ask students what happens to this equation when we plug in $h = 0$, I always get a chorus of "undefined!" This shows good mathematical training — division by zero is indeed undefined — but doesn't give any real insight into what will go wrong when the workers try to carry out the mayor's plan. If we make h smaller and smaller T will get bigger and bigger. By making h sufficiently small, we can make T arbitrarily large. The important insight here is that a quantity like $1/0$ isn't just undefined, it's undefined because it's infinity, and infinity isn't a real number. If the workers actually try to make $h = 0$, they will simply have to tighten the wires so much that the wires break.

Another way of putting this is that the limit $\lim_{h\to 0} T(h)$ fails to exist. Although it's true that the limit doesn't exist, we can be more descriptive about the reason that it doesn't. It's a limit that doesn't exist because it equals infinity.

[1] The value of k is $WL/8$, where W is the weight of the wire and L is the horizontal length. The approximation is good if h is small compared to L.



e / The function $1/x$ behaves badly near $x = 0$.

Consider the limit

$$\lim_{x\to 0}\frac{1}{x}.$$

As x decreases to $x = 0$ through smaller and smaller positive values, its reciprocal $1/x$ becomes larger and larger. We say that instead of going to some finite number, the quantity $1/x$ "goes to infinity" as $x \searrow 0$. In symbols:

$$\lim_{x\searrow 0}\frac{1}{x} = \infty. \tag{5}$$

Likewise, as x approaches 0 through negative numbers, its reciprocal $1/x$ drops lower and lower, and we say that $1/x$ "goes to $-\infty$" as $x \nearrow 0$. Symbolically,

$$\lim_{x\nearrow 0}\frac{1}{x} = -\infty. \tag{6}$$

The limits (5) and (6) are not like the normal limits we have been dealing with so far. Namely, when we write something like

$$\lim_{x\to 2} x^2 = 4$$

we mean that the limit actually exists and that it is equal to 4. On the other hand, since we have agreed that ∞ is not a number (see p. 65), the meaning of (5) cannot be to say that "the limit exists and its value is ∞."

Instead, when we write

$$\lim_{x\to a} f(x) = \infty \tag{7}$$

for some function $y = f(x)$, we mean, *by definition,* that the limit of $f(x)$ does not exist, and that it fails to exist in a specific way: as $x \to a$, the value of $f(x)$ becomes "larger and larger," and in fact eventually becomes larger than any finite number.

The language in that last paragraph shows you that this is an intuitive definition, at the same level as the first definition of limit we gave in section 2.1.1, p. 48. It contains the usual suspect phrases such as "larger and larger," or "finite number" (as if there were any other kind.) A more precise definition involving epsilons can be given, but in this course we will not go into this much detail.

When a function is going to blow up at a certain point, there are two common behaviors. The first is the one shown in figure e for $1/x$, where the limit is $+\infty$ on one side and $-\infty$ on the other. If a limit is to be more than a one-sided limit, we want it to have the same value on the left and right. In this example that doesn't happen, so only the one-sided limits can be described as being positive- or negative-infinite:

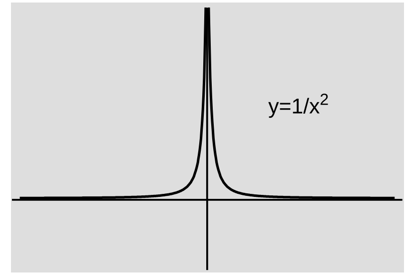
$$\lim_{x\searrow 0}\frac{1}{x} = +\infty$$

$$\lim_{x\nearrow 0}\frac{1}{x} = -\infty$$

$$\lim_{x\to 0}\frac{1}{x} \text{ can't be described as } +\infty \text{ or } -\infty$$

The function $1/x^2$, figure f, exhibits the other frequently encountered behavior. Here we have a positive blowup on both sides, so it isn't just the one-sided limits that can be described.

$$\lim_{x \searrow 0} \frac{1}{x^2} = +\infty$$
$$\lim_{x \nearrow 0} \frac{1}{x^2} = +\infty$$
$$\lim_{x \to 0} \frac{1}{x^2} = +\infty$$



f / The function $1/x^2$ blows up near $x = 0$, but in a different way than $1/x$; it approaches *positive* infinity on both sides.

As a final comment on infinite limits, it is important to realize that (7) is not a normal limit, and *you cannot apply the limit rules to infinite limits.* Here is an example of what goes wrong if you try anyway.

Trouble with infinite limits Example 9

If you apply the limit properties to $\lim_{x\searrow 0} 1/x = \infty$, then you could conclude

$$1 = \lim_{x\searrow 0} x \cdot \frac{1}{x} = \lim_{x\searrow 0} x \times \lim_{x\searrow 0} \frac{1}{x} = 0 \times \infty = 0,$$

because "anything multiplied with zero is zero."

After using the limit properties in combination with this infinite limit we reach the absurd conclusion that 1 = 0. The moral of this story is that you can't use the limit properties when some of the limits are infinite.

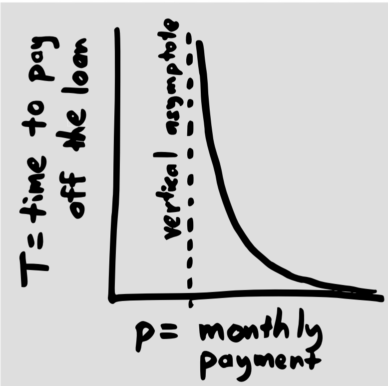
4.4 Curve sketching

4.4.1 Sketching a graph without knowing its equation

The concepts of calculus, such as derivatives, limits, curvature, and concavity, can guide us in analyzing the behavior of a function even when we don't know a formula for the function. In economics, for example, these concepts are used heavily even though real-world data can essentially never be described by a formula. This subsection presents four examples in which we can use these concepts to sketch a function based on our understanding of how the function should behave in real life.

The time to pay off a loan

Most people will end up borrowing money at some point in their lives, whether it's credit card debt, a mortgage, a loan to buy a car, or a cash advance from a payday loan company. One of the warning signs that you may be walking into an exploitative situation is if the person trying to sell you the loan emphasizes the low monthly payment. Suppose that you're borrowing \$10,000 to buy a car, and the monthly interest rate is 1%. Let p be the monthly payment, and T the time required in order to pay off the loan. To understand what's going on here, you want to be able to *visualize* the graph of T as a function of p. One fairly tedious way to do this would be to find the equation of the function, take a piece of graph paper and plot points. Another method would be to use an expensive graphing calculator. But your knowledge of calculus gives you a method that provides more insight with less work.



g / The time required to pay off a loan, as a function of the monthly payment.

Clearly the smaller the payment, the longer it will take to pay off the loan. This tells us that $T(p)$ is a *decreasing* function; its derivative will always be negative.

If p is large, then you will pay off the loan so quickly that no significant amount of interest accrues. Therefore at large values of p, we will have $T \approx (\$10,000)/p$. This tells us that $\lim_{p\rightarrow\infty} T = 0$. The graph of T will approach the horizontal axis more and more closely as p gets bigger and bigger. We say that the function $T(p)$ has a *horizontal asymptote* at zero.

Finally, what happens if p is small? Remember, interest on the loan is accruing at a rate of 1% monthly, or \$100 every month. It may sound like a good deal if you're offered this loan with a low monthly payment of \$101, but if you take the loan and always make the minimum payment, then the principal on the loan will only go down by \$1 every month. You will die of old age before you pay off the car. We can therefore tell that $\lim_{p\searrow\$100} T = \infty$. This is a *vertical asymptote* on the graph.

Figure g shows what the graph must look like.

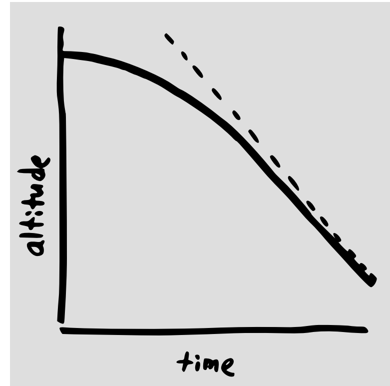
The Laffer curve

This example, a famous one, also has to do with money. In 1974, economist Arthur Laffer presented the following argument about taxes to politicians Dick Cheney and Donald Rumsfeld, sketching the resulting graph on a paper napkin. Consider the government's tax revenue as a function of the tax rate. Clearly if the tax rate is zero, the government gets zero revenue. Most people would assume that the function was a purely increasing one, since raising the tax rate would always garner the government more money.

h / The Laffer curve.

But, Laffer said, that isn't so. Imagine that the tax rate was 100%, so that the government confiscated all of everyone's earnings. Nobody would have any incentive to work, so they would stop working, they would earn no taxable income, and revenue would drop to zero. Laffer sketched a graph like figure h on a paper napkin for Cheney and Rumsfeld. There should be some intermediate tax rate, he told them, that would produce the maximum revenue. Later, when Ronald Reagan became president, he cut taxes on the theory that the US was already on the right-hand side of the "Laffer curve," so that, counterintuitively, the lower taxes would produce *higher* revenue. The results were not as Laffer had promised; the average annual budget deficit during the Reagan administration was \$240 billion, compared to \$57 billion during the preceding Carter administration.

In calculus terms, our analysis of this function is an example of a result called Rolle's theorem, p. 117. The idea is that if the function is smooth, then we expect its derivative to be continuous. If the derivative is positive on the left and negative on the right, then it must be zero at some intermediate point. This would be the point at which the function was maximized.



i / Altitude as a function of time for a skydiver.

Skydiving

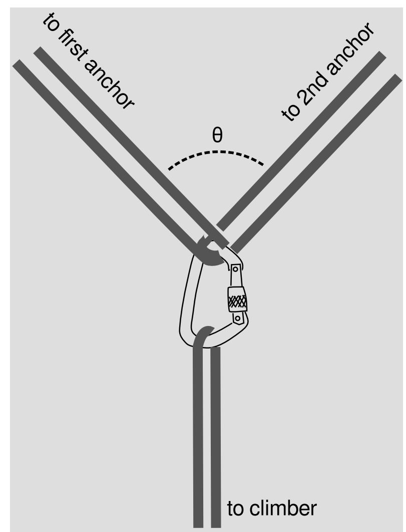
Figure i shows a skydiver's altitude as a function of time. Early in the motion, soon after the person jumps out of the plane, the only significant force is gravity, and the person falls with constant acceleration (section 1.5.1, p. 22). The drop relative to the initial position equals $(1/2)at^2$, which is the equation of a parabola.

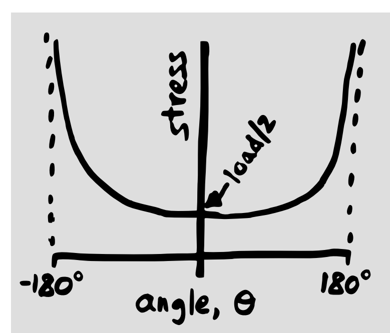
But as the downward (negative) velocity increases, the upward force of air friction gets stronger and stronger. In the opposite limit of $t \to \infty$, the force of air friction gets closer and closer to being strong enough to cancel the force of gravity. In this limit, Newton's second law (section 3.4.2, p. 88) predicts an acceleration of zero. An acceleration of zero corresponds to constant velocity, so that the graph asymptotically approaches a line whose slope is the velocity.

This graph demonstrates two mathematical properties. It has a *y-intercept*, which is the initial altitude. It also has an oblique asymptote, i.e., an asymptotic line that is neither horizontal nor vertical.

A rock-climbing anchor

For safety, rock climbers and mountaineers often wear a climbing harness and tie in to other climbers on a rope team or to anchors such as pitons or snow anchors. When using anchors, the climber usually wants to be protected by more than one, both for extra strength and for redundancy in case one fails. Figure j shows such an arrangement, with the climber hanging from a pair of anchors forming a "Y" at an angle θ. The usual advice is to make $\theta < 90°$; for large values of θ, the stress placed on the anchors can be many times greater than the actual load L, so that two anchors are actually *less* safe than one.





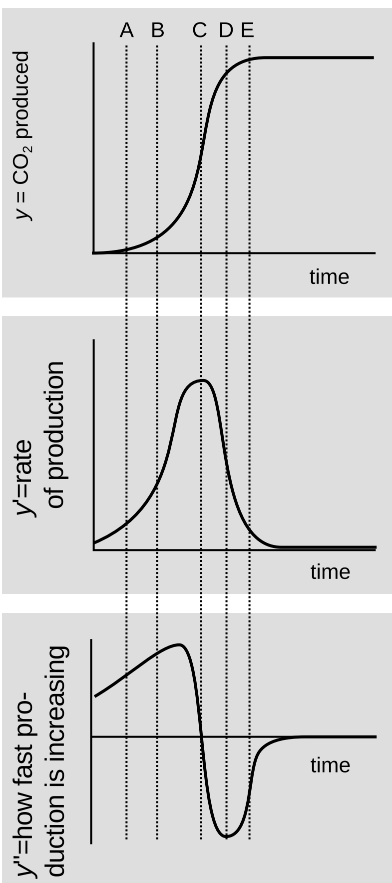
j / A rock-climbing anchor.

Consider the stress on the anchor S as a function of θ. For physical reasons similar to those discussed in the example of the telephone wire (section 4.3.3, p. 103), S must approach infinity as θ approaches 180 degrees; no matter how tight the anchor strands are made, the carabiner (hook) at the center will never be pulled up quite as high as the anchors.

At $\theta = 0$, we can see that each anchor strand will support half the load. The *y-intercept* of the graph equals $L/2$.

We can gain further insight by extending the range of possible values for θ to include negative angles. Physically, this corresponds to bringing the anchor strands past one another and swapping the roles of the two anchors. Since the physical setup is symmetrical, the function $S(\theta)$ must have the property $S(\theta) = S(-\theta)$, i.e., it is an *even* function. It might seem pointless to discuss this symmetry, but it tells us something important. An argument identical to the one in section 1.2.4, p. 17, tells us that based on this symmetry, the derivative S' must equal zero at $\theta = 0$. This means that for small values of θ, the strain on the anchor will be very nearly the same as for $\theta = 0$, i.e., hardly any greater than half the load. Thus any small value of θ is about equally good, but very large values could be a deadly mistake.

4.4.2 Sketching f' and f'' given the graph of f



k / Sketching y' and y'' given the graph of y.

In figure k we revisit the example of fermenting beer (section 3.1, p. 83). (Feel free to mark your place in the book and make a trip to the fridge before continuing.) The top panel of the graph would probably have been the easiest to sketch starting from scratch. Clearly the amount of CO_2 produced starts off at zero, it rises, and it must eventually flatten out and approach a horizontal asymptote, since the yeast use up all their food and can't produce any more. This kind of vaguely S-shaped curve is in fact encountered in many situations, and is often referred to as a "yeast curve."

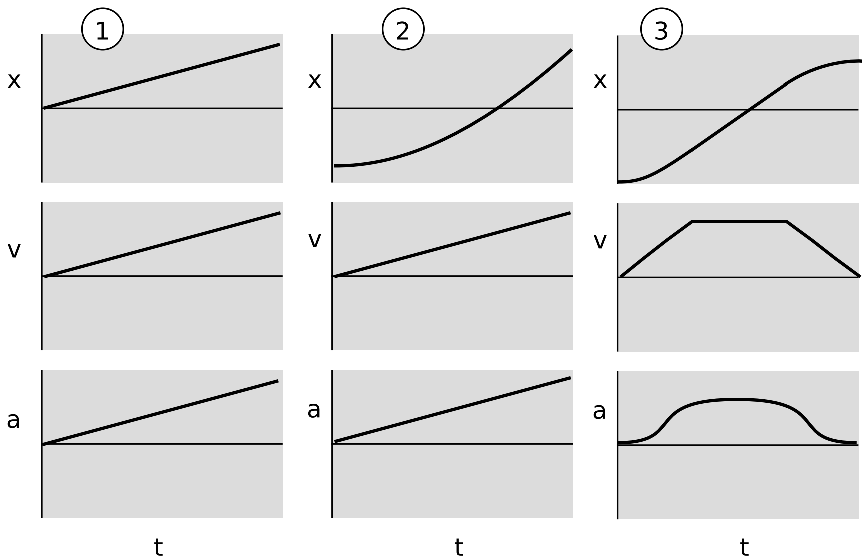
Now suppose we know y and we want to find y' and y''. The basic concept is that the *slope* of each graph in the stack gives the *value* of the graph below it. The slope of the tangent line to the y graph at time A is small and positive, while the slope at B is larger and positive. Therefore the *values* of y' at these times must be small and positive, then larger and positive. At time C, the slope of the y graph is as great as it will ever be. Therefore the y' graph has a maximum there. The slope of y gets smaller at D and still smaller at E, so the value of y' must taper off correspondingly.

Now that we've sketched the graph of y', we can continue the process and construct *its* derivative, y''. At time C the slope of the y' graph is zero, so the value of the y'' graph is zero; this is a point of inflection. At times earlier than C the slope of y' is positive, while at times later than C it's negative. Therefore we must have $y'' > 0$ before C and $y'' < 0$ after.

We can also relate the properties of the y'' graph directly to those of the y graph. The second derivative is a measure of curvature, and its sign indicates concavity. The y graph is concave up before C and concave down after. This matches up with the signs of y''.

Discussion question

A Figure I shows three stacks of graphs, each of which is supposed to represent the position, velocity, and acceleration of an object. Explain how each set of graphs contains inconsistencies, and fix them.



I / Discussion question A.

4.4.3 Sketching a graph given its equation

If we have an equation defining a function, then the following procedure is often a fairly efficient way of sketching its graph. Often we are especially interested in finding the function's local maxima and minima, including the *absolute* or *global* maxima and minima. That is, the absolute maximum is the greatest value ever attained by the function, and similarly for the absolute minimum.

1. Find all solutions of $f'(x) = 0$ in the interval $[a, b]$: these are called the *critical* or *stationary* points for f.
2. Find the sign of $f'(x)$ at all other points.
3. Each stationary point at which $f'(x)$ actually changes sign is a local maximum or local minimum. Compute $f(x)$ at each stationary point.
4. Compute the values of the function $f(a)$ and $f(b)$ at the endpoints of the interval.
5. The absolute maximum is attained at the stationary point or the boundary point with the highest value of f; the absolute minimum occurs at the boundary or stationary point with the smallest value.

If the interval is unbounded, then instead of computing the values $f(a)$ or $f(b)$, you should instead compute $\lim_{x\to\pm\infty} f(x)$.

As an example, let's sketch the graph of the rational function

$$f(x) = \frac{x(3-4x)}{1+x^2}.$$

By looking at the signs of numerator and denominator we see that

$$f(x) > 0 \text{ for } 0 < x < \tfrac{3}{4}$$
$$f(x) < 0 \text{ for } x < 0 \text{ and also for } x > \tfrac{3}{4}.$$

We compute the derivative of f:

$$f'(x) = \frac{-3x^2 - 8x + 3}{(1+x^2)^2}.$$

Hence $f'(x) = 0$ if and only if

$$-3x^2 - 8x + 3 = 0,$$

and the solutions to this quadratic equation are -3 and $1/3$. These two roots will appear several times, and it will shorten our formulas if we abbreviate

$$A = -3 \text{ and } B = 1/3.$$

To see if the derivative changes sign we factor the numerator and denominator. The denominator is always positive, and the numerator is

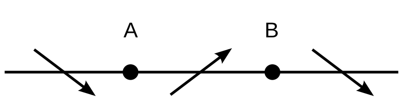
$$-3x^2 - 8x + 3 = -3\left(x^2 + \frac{8}{3}x - 1\right) = -3(x-A)(x-B).$$

Therefore

$$f'(x) \begin{cases} < 0 & \text{for } x < A \\ > 0 & \text{for } A < x < B \\ < 0 & \text{for } x > B \end{cases}$$

It follows that f is decreasing on the interval $(-\infty, A)$, increasing on the interval (A, B) and decreasing again on the interval (B, ∞) (figure m). Therefore

$$A \text{ is a local minimum, and } B \text{ is a local maximum.}$$



m / The sign of the derivative changes at A and B.

Are these global maxima and minima?

Since we are dealing with an unbounded interval we must compute the limits of $f(x)$ as $x \to \pm\infty$. We find

$$\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = -4.$$

Since f is decreasing between $-\infty$ and A, it follows that

$$f(A) \le f(x) < -4 \text{ for } -\infty < x \le A.$$

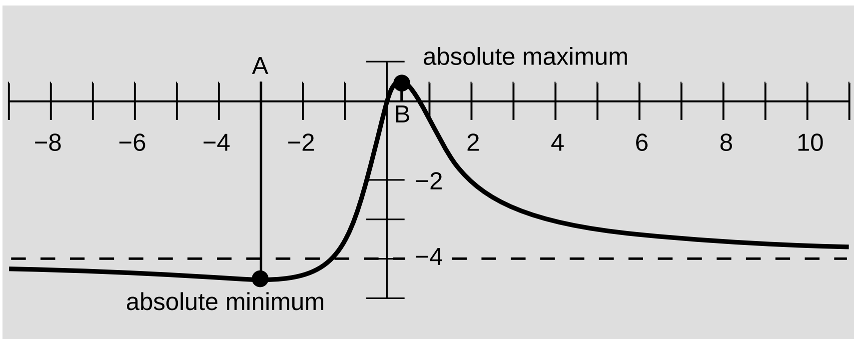
Similarly, f is decreasing from B to $+\infty$, so

$$-4 < f(x) \leq f(B) \text{ for } B < x < \infty.$$

Between the two stationary points the function is increasing, so

$$f(A) \leq f(x) \leq f(B) \text{ for } A \leq x \leq B.$$

From this it follows that $f(x)$ has a global minimum when $x = A = -3$ and has a global maximum when $x = B = 1/3$.

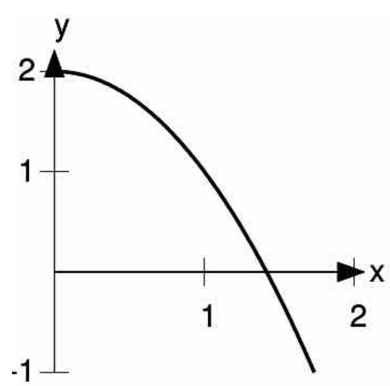


n / The graph of $f(x) = x(3 - 4x)/(1 + x^2)$.

4.5 Completeness

4.5.1 The completeness axiom of the real numbers

Calculus is the study of rates of change (differentiation) and how change accumulates (integration, which we haven't encountered yet). What changes is always a *function*, and the function takes an input value that belongs to its domain and gives back an output that belongs to its range. The domain and range could in principle be sets of integers, rational numbers, real numbers, complex numbers, or hyperreal numbers (section 2.9, p. 64). These number systems all share many of the same properties, but just as the ocean is the natural setting for a pirate story, there is a sense in which the real numbers are the natural setting in which to do calculus. Throughout this book, without specifically commenting on it so far, we've been considering only functions that take real-number inputs and give back real-number outputs: real functions.



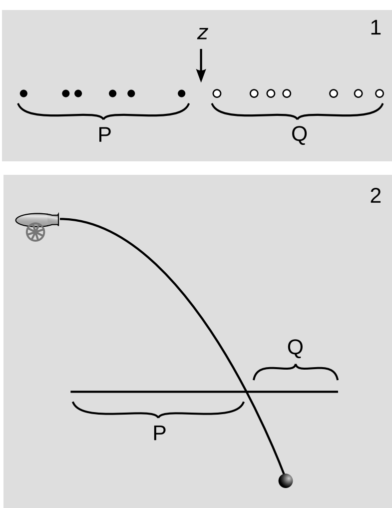
o / A cannonball is fired horizontally, and hits the water at $y = 0$.

What's so special about real functions? We can define functions whose inputs and outputs are, say, integers, and such functions are of interest in many fields of mathematics. But real functions are especially well suited to describing rates of change. As an example, the graph in figure o shows the function $f(x) = 2 - x^2$. Let's say this represents the arc of a cannon-ball shot off of a cliff into the ocean, where a y coordinate of 0 represents the surface of the water. Our geometrical intuition tells us that if the ball starts above the water, and later on ends up below it, then there must be some point at which it enters the water. In other words, if the graph of the function f cuts across the line $y = 0$, then there must be a point at which they coincide.

But if we consider a set of numbers more restricted than the real numbers, this may not happen. For example, suppose we take f to be a function whose inputs and outputs are rational numbers. Recall that a rational number is any number that can be expressed as an integer divided by another integer, e.g., the fraction 2/3. But the place where our cannonball crosses sea level has $x = \sqrt{2}$, which is not a rational number. This example shows that the graphs of two rational-number functions can cut across one another without ever touching! This offends our intuition about rates of change, since we expect that if we change a variable smoothly from one value to another, it should visit every value in between.

What is the special ingredient, the secret sauce that allows the real number system to avoid such paradoxical results as the one about the cannonball? It seems that the reals are somehow more *densely packed* on the number line than the rationals, but how do we define this density property in mathematical terms? It can't be any of the elementary properties of the reals (section 1.6, p. 25), since the rationals also satisfy all of those properties. We need to add a new axiom, which is called the completeness axiom.

One possible way of stating such an axiom is the following.



p / 1. The sets P and Q are separated on the number line so that every point in P is to the left of every point in Q. By the completeness axiom, a number like z exists. 2. By the completeness axiom, the curve $f(x) = 2 - x^2$ must intersect the axis. The point of intersection is $z = \sqrt{2}$. The completeness axiom doesn't hold for the rational numbers, and we can see that here because z is an irrational number.

Completeness axiom
Let P and Q be sets of numbers such that every number in P is smaller than every number in Q. Then there exists some number z such that z is greater than or equal to every number in P, but less than or equal to any number in Q.

As an example, let P be the set of all numbers x such that $x^2 < 2$, and Q the set of x such that $x^2 \geq 2$. Then the number z would have to be $\sqrt{2}$, which shows that the rationals are not complete. The reals are complete, and the completeness axiom can serve as one of the fundamental axioms of the real numbers.

The completeness axiom is of a fundamentally different character than the elementary axioms. The elementary axioms make statements such as "for any number x, ..." or "for any numbers x and y, ..." The completeness axiom says "for any *sets* of numbers P and Q, ..."

Every decimal is a real number Example 10

Consider the infinite decimal

$$3.141592\ldots,$$

which is the decimal expansion of π. We can use the completeness axiom to prove that this is a real number. Let P be the list of rational numbers given by $\{3,\ 3.1,\ 3.14,\ 3.141,\ \ldots\}$. Let Q be the set of rational numbers that are larger than every number in P. Then the real number whose existence is asserted by the completeness axiom is exactly π. Similar reasoning shows that any decimal corresponds to some real number (which can be shown to be unique). (Note, however, that the same real number can have more than one decimal expansion. For example, the infinite repeating decimals $1.000\ldots$ and $0.999\ldots$ both equal 1.)

The Archimedean property Example 11

The Archimedean principle states that there is no positive real number that is less than $1/1$, less than $1/(1+1)$, less than $1/(1+1+1)$, and so on.[2] In other words, it says that there are no real numbers that are infinitely small, but still greater than zero. The Archimedean property can be proved from the completeness property. For suppose, to the contrary, that we did have such a real number. Then it would be less than $1/10$, so its first decimal place would be 0. It would also be less than $1/100$, so its second decimal place would also be zero. Continuing in this way, we find that the decimal expansion of such a number must be $0.000\ldots$, with the zeroes repeating forever. But this is the decimal expansion of zero, and we already know that every decimal expansion corresponds to a unique real number. Therefore our number is zero, and this is a contradiction, since we assumed that it violated the Archimedean principle, which refers to a *positive* real number.

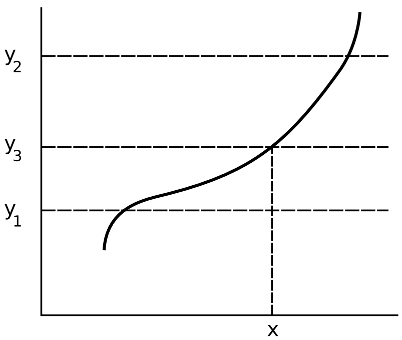
[2] Cf. section 2.9, p. 64. For an application to economics, see rule 3, p. 218.

4.5.2 The intermediate and extreme value theorems

The following two theorems can be proved from the completeness property and the elementary properties of the reals, but we will not give the proofs here.

The intermediate value theorem

Intuitively, the intermediate value theorem says that the real numbers aren't susceptible to paradoxes like the cannonball paradox described above. Or, we can say that if you are moving continuously along a road, and you get from point A to point B, then you must also visit every other point along the road; only by teleporting (by moving discontinuously) could you avoid doing so. More formally, the theorem says this:



q / The intermediate value theorem states that if the function is continuous, it must pass through y_3.

Intermediate value theorem
If y is a continuous real-valued function on the real interval from a to b, and if y takes on values y_1 and y_2 at certain points within this interval, then for any y_3 between y_1 and y_2, there is some real x in the interval for which $y(x) = y_3$.

Example 12

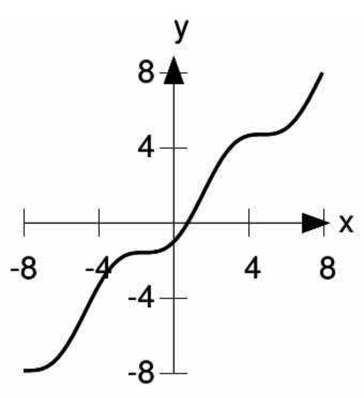
▷ Show that there is a solution to the equation $10^x + x = 1000$.

▷ We expect there to be a solution near $x = 3$, where the function $f(x) = 10^x + x = 1003$ is just a little too big. On the other hand, $f(2) = 102$ is much too small. Since f has values above and below 1000 on the interval from 2 to 3, and f is continuous, the intermediate value theorem proves that a solution exists between 2 and 3. If we wanted to find a better numerical approximation to the solution, we could do it using Newton's method, which is introduced in section 7.2.

Example 13

▷ Show that there is at least one solution to the equation $\cos x = x$, and give bounds on its location.

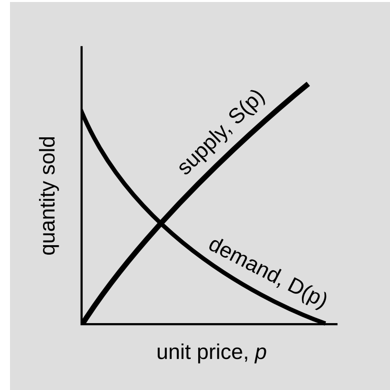
▷ This is what's known as a transcendental equation, and no amount of fiddling with algebra and trig identities will ever give a closed-form solution, i.e., one that can be written down with a finite number of arithmetic operations to give an exact result. However, we can easily prove that at least one solution exists, by applying the intermediate value theorem to the function $f(x) = x - \cos x$. The cosine function is bounded between -1 and 1, so f must be negative for $x < -1$ and positive for $x > 1$. By the intermediate value theorem, there must be a solution in the interval $-1 \le x \le 1$. The graph, r, verifies this, and shows that there is only one solution.



r / The function $x - \cos x$ constructed in example 13.

Supply and demand Example 14

Figure s shows two graphs representing the supply and demand of some good on a free market. The function $D(p)$ shows the quantity that buyers would willingly buy at unit price p. Normally D is a decreasing function: if the price goes up, people don't buy as much. (But cf. problem c4, p. 37.) The function $S(p)$ shows the quantity that the seller would willingly offer if the unit price was p. Often S is an increasing function. For example, Boeing might only be able to produce more passenger jets by paying their workers overtime, which would create a cost that they would pass on to their customers.



s / Example 14.

Suppose that, as in the example shown in the figure, D starts out higher than S on the left, but ends up lower than S on the right. Then we expect geometrically that if the curves are continuous, they must cross at some point. This can be proved using the same technique as in example 13. We construct a function $f(p) = S(p) - D(p)$, which goes from negative to positive. By the intermediate value theorem, there must be some point where $f = 0$, meaning that $S = D$. This crossing point is the free-market equilibrium.

The intermediate value theorem holds for real numbers, but in fact neither the price nor the quantity is free to have any real-number value. For example, Boeing can't sell half an airplane. In some cases this might mean that the free-market equilibrium defined by $S = D$ would not exist. An example might be the Concorde, a supersonic passenger jet, which flew from 1969 to 2003. The nonexistence of the market for this plane today may indicate that the supply and demand curves now cross at a quantity that is greater than 0 and less than 1, which is not a possible free-market equilibrium because the planes can only be sold in whole numbers.

Example 15

▷ Prove that every odd-order polynomial P with real coefficients has at least one real root x, i.e., a point at which $P(x) = 0$.

▷ Example 13 might have given the impression that there was nothing to be learned from the intermediate value theorem that couldn't be determined by graphing, but this example clearly can't be solved by graphing, because we're trying to prove a general result for all polynomials.

To see that the restriction to odd orders is necessary, consider the polynomial x^2+1, which has no real roots because $x^2 > 0$ for any real number x.

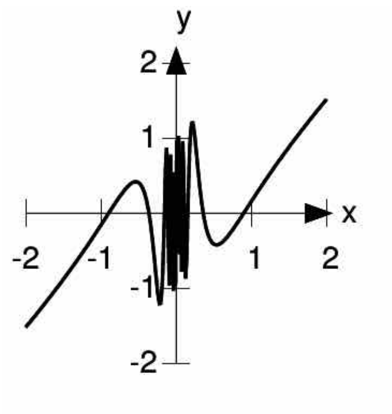
To fix our minds on a concrete example for the odd case, consider the polynomial $P(x) = x^3 - x + 17$. For large values of x, the linear and constant terms will be negligible compared to the x^3

term, and since x^3 is positive for large values of x and negative for large negative ones, it follows that P is sometimes positive and sometimes negative. Therefore by the intermediate value theorem P has at least one root.

This argument didn't depend much on the specific polynomial P chosen as an example. The fact that P was positive for large x and negative for large negative x followed merely from the fact that P was of odd order. Therefore the result holds for all polynomials of odd order.

Example 16

▷ Show that the equation $x = \sin 1/x$ has infinitely many solutions.



t / The function $x - \sin 1/x$.

▷ This is another example that can't be solved by graphing; there is clearly no way to prove, just by looking at a graph like t, that the function $f(x) = x - \sin 1/x$ crosses the x axis *infinitely* many times. The graph does, however, help us to gain intuition for what's going on. As x gets smaller and smaller, $1/x$ blows up, and $\sin 1/x$ oscillates more and more rapidly. The function f is undefined at 0, but it's continuous everywhere else, so we can apply the intermediate value theorem to any interval that doesn't include 0.

We want to prove that for any positive u, there exists an x with $0 < x < u$ for which $f(x)$ has either desired sign. Let n be an even integer such that $n > 10$ and also $\pi n > 1/u$. Then clearly $f(x)$ is negative at $x = 1/(\pi n + \pi/2) < u$, since $\sin 1/x = 1$ and x is small. Similarly, $f(x)$ is positive at $x = 1/(\pi n + 3\pi/2) < u$. This establishes the desired result.

The extreme value theorem

We've seen that that locating maxima and minima of functions may in general be fairly difficult, because there are so many different ways in which a function can attain an extremum: e.g., at an endpoint, at a place where its derivative is zero, or at a nondifferentiable kink. The following theorem allows us to make a very general statement about all these possible cases, assuming only continuity.

> **Extreme value theorem**
> If f is a continuous real-valued function on the real-number interval defined by $a \le x \le b$, then f has maximum and minimum values on that interval, which are attained at specific points in the interval.

Let's first see why the assumptions are necessary. If we weren't confined to a finite interval, then $y = x$ would be a counterexample, because it's continuous and doesn't have any maximum or minimum value. If we didn't assume continuity, then we could have a function defined as $y = x$ for $x < 1$, and $y = 0$ for $x \ge 1$; this function never gets bigger than 1, but it never attains a value of 1 for any specific value of x. If we didn't assume a real function, then we could have, for example, the function $f(x) = (x^2 - 2)^2$ defined on the rational numbers, which would never attain the minimum value of 0 because $\sqrt{2}$ isn't a rational number.

▷ Example 17

Find the maximum value of the polynomial $P(x) = x^3 + x^2 + x + 1$ for $-5 \leq x \leq 5$.

▷ Polynomials are continuous, so the extreme value theorem guarantees that such a maximum exists. Suppose we try to find it by looking for a place where the derivative is zero. The derivative is $3x^2 + 2x + 1$, and setting it equal to zero gives a quadratic equation, but application of the quadratic formula shows that it has no real solutions. It appears that the function doesn't have a maximum anywhere (even outside the interval of interest) that looks like a smooth peak. Since it doesn't have kinks or discontinuities, there is only one other type of maximum it could have, which is a maximum at one of its endpoints. Plugging in the limits, we find $P(-5) = -104$ and $P(5) = 156$, so we conclude that the maximum value on this interval is 156.

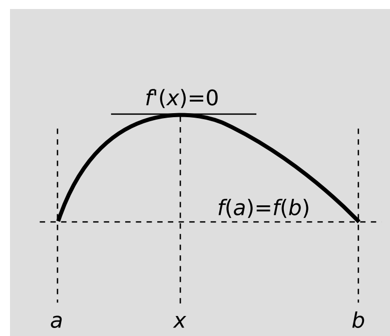
4.5.3 Rolle's theorem and the mean-value theorem

On p. 106, in the example of the Laffer curve from economics, we got a preview of the following intuitively appealing theorem.

Rolle's theorem

Let f be a function that is continuous on the interval $[a, b]$ and differentiable on (a, b), and let $f(a) = f(b)$. There there exists a point $x \in (a, b)$ such that $f'(x) = 0$.

Proof: By the extreme value theorem, f attains its maximum and minimum values in $[a, b]$. If both of these are at endpoints, then f is a constant function, and the theorem holds trivially. Suppose instead that at least one of these extrema is on the interior of the interval. Then by the theorem given in section 2.8.3, f' is zero at that point, and the theorem also holds.□



u / Rolle's theorem.

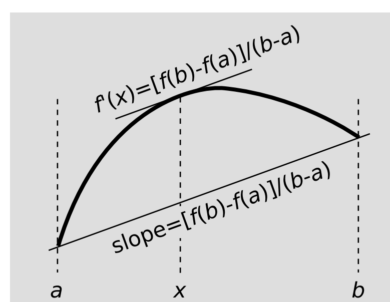
Rolle's theorem can be straightforwardly generalized to the following.

Mean value theorem

Let f be a function that is continuous on the interval $[a, b]$ and differentiable on (a, b). There there exists a point $x \in (a, b)$ such that

$$f'(x) = \frac{f(b) - f(a)}{b - a},$$

meaning that the derivative equals the average (mean) rate of change of the function between the endpoints of the interval.



v / The mean value theorem.

"Mean" is just a fancy word for "average." In general, it's a mistake to try to calculate a rate of change without calculus, using $\Delta y/\Delta x$, unless the rate of change is constant. The mean value theorem says that just as a broken clock is right twice a day, there is at least one point where $\Delta y/\Delta x$ gives the right answer.

Proof: Define the function

$$\ell(x) = a + \frac{f(b) - f(a)}{b - a}(x - a),$$

which is the point-slope form of the line passing through the end-points of the graph of f. Define a new function $g(x) = f(x) - \ell(x)$, so that $g(a) = g(b) = 0$. Applying Rolle's theorem to g, we find that there is some point where $f'(x) = \ell'(x)$, which is the desired result.□

4.6 Two tricks with limits

4.6.1 Rational functions that give 0/0

Suppose we want to compute the following limit:

$$\lim_{x\to 2} \frac{x^2 - 2x}{x^2 - 4}$$

We first use the limit properties to find

$$\lim_{x\to 2} x^2 - 2x = 0 \text{ and } \lim_{x\to 2} x^2 - 4 = 0.$$

Now to complete the computation we would like to apply the property (P_6) about quotients, but this would give us

$$\lim_{x\to 2} f(x) = \frac{0}{0}.$$

The denominator is zero, so we were not allowed to use (P_6) (and the result doesn't mean anything anyway). We have to do something else.

The function we are dealing with is a *rational function*, which means, as mentioned in example 7, p. 102, that it is the quotient of two polynomials. For such functions there is an algebra trick that always allows you to compute the limit even if you first get $\frac{0}{0}$. The thing to do is to divide numerator and denominator by $x - 2$. In our case we have

$$x^2 - 2x = (x - 2) \cdot x, \qquad x^2 - 4 = (x - 2) \cdot (x + 2)$$

so that

$$\lim_{x\to 2} f(x) = \lim_{x\to 2} \frac{(x - 2) \cdot x}{(x - 2) \cdot (x + 2)} = \lim_{x\to 2} \frac{x}{x + 2}.$$

After this simplification we *can* use the properties $(P_{\dots})$ to compute

$$\lim_{x\to 2} f(x) = \frac{2}{2 + 2} = \frac{1}{2}.$$

4.6.2 The "don't make δ too big" trick

In this section we describe a trick, the "don't make δ to too big" trick, that is sometimes helpful when we want to evaluate a limit directly from the epsilon-delta definition. Say we want to prove that $\lim_{x\to 1} x^2 = 1$. This may not seem to require a fancy proof, since obviously plugging in $x = 1$ gives $x^2 = 1$. But since functions can be discontinuous, plugging in does not always prove the value of a limit. Also, this example will be an excuse to develop a technique that can be useful in less trivial cases.

We have $f(x) = x^2$, $a = 1$, $L = 1$, and as usual when computing a limit the question is, "how small should $|x - 1|$ be to guarantee $|x^2 - 1| < \varepsilon$?"

We begin by estimating the difference $|x^2 - 1|$

$$|x^2 - 1| = |(x - 1)(x + 1)| = |x + 1| \cdot |x - 1|.$$

As x approaches 1 the factor $|x-1|$ becomes small, and if the other factor $|x + 1|$ were a constant (e.g. 2 as in the previous example) then we could find δ as before, by dividing ε by that constant.

Here is a trick that allows you to replace the factor $|x+1|$ with a constant. We hereby agree *that we always choose our δ so that $\delta \le 1$.* If we do that, then we will always have

$$|x - 1| < \delta \le 1, \text{i.e. } |x - 1| < 1,$$

and x will always be between 0 and 2. Therefore

$$|x^2 - 1| = |x + 1| \cdot |x - 1| < 3|x - 1|.$$

If we now want to be sure that $|x^2 - 1| < \varepsilon$, then this calculation shows that we should require $3|x - 1| < \varepsilon$, i.e. $|x - 1| < \frac{1}{3}\varepsilon$. So we should choose $\delta \le \frac{1}{3}\varepsilon$. We must also live up to our promise never to choose $\delta > 1$, so if we are handed an ε for which $\frac{1}{3}\varepsilon > 1$, then we choose $\delta = 1$ instead of $\delta = \frac{1}{3}\varepsilon$. To summarize, we are going to choose

$$\delta = \text{the smaller of 1 and } \frac{1}{3}\varepsilon.$$

We have shown that if you choose δ this way, then $|x-1| < \delta$ implies $|x^2 - 1| < \varepsilon$, no matter what $\varepsilon > 0$ is.

The expression "the smaller of a and b" shows up often, and is abbreviated to $\min(a, b)$. We could therefore say that in this problem we will choose δ to be

$$\delta = \min\bigl(1, \tfrac{1}{3}\varepsilon\bigr).$$

Example 18

▷ Show that $\lim_{x\to 4} 1/x = 1/4$.

▷ We apply the definition with $a = 4$, $L = 1/4$ and $f(x) = 1/x$. Thus, for any $\varepsilon > 0$ we try to show that if $|x-4|$ is small enough then one has $|f(x) - 1/4| < \varepsilon$.

We begin by estimating $|f(x) - \frac{1}{4}|$ in terms of $|x-4|$:

$$|f(x) - 1/4| = \left|\frac{1}{x} - \frac{1}{4}\right| = \left|\frac{4-x}{4x}\right| = \frac{|x-4|}{|4x|} = \frac{1}{|4x|}\,|x-4|.$$

As before, things would be easier if $1/|4x|$ were a constant. To achieve that we again agree not to take $\delta > 1$. If we always have $\delta \le 1$, then we will always have $|x-4| < 1$, and hence $3 < x < 5$. How large can $1/|4x|$ be in this situation? Answer: the quantity $1/|4x|$ increases as you decrease x, so if $3 < x < 5$ then it will never be larger than $1/|4 \cdot 3| = \frac{1}{12}$.

We see that if we never choose $\delta > 1$, we will always have

$$|f(x) - \tfrac{1}{4}| \le \tfrac{1}{12}|x-4| \quad \text{for} \quad |x-4| < \delta.$$

To guarantee that $|f(x) - \frac{1}{4}| < \varepsilon$ we could therefore require

$$\tfrac{1}{12}|x-4| < \varepsilon, \quad \text{i.e.} \quad |x-4| < 12\varepsilon.$$

Hence if we choose $\delta = 12\varepsilon$ or any smaller number, then $|x-4| < \delta$ implies $|f(x) - 4| < \varepsilon$. Of course we have to honor our agreement never to choose $\delta > 1$, so our choice of δ is

$$\delta = \text{the smaller of 1 and } 12\varepsilon = \min\bigl(1, 12\varepsilon\bigr).$$

Problems

a1 Suppose x is a big, positive number. Experiment on a calculator to figure out whether $\sqrt{x+1}-\sqrt{x-1}$ comes out big, normal, or tiny. Try making x bigger and bigger, and see if you observe a trend. Based on these numerical examples, form a conjecture about the limit of this expression as x approaches infinity.

▷ Solution, p. 232

a2 If we want to pump air or water through a pipe, common sense tells us that it will be easier to move a larger quantity more quickly through a fatter pipe. Quantitatively, we can define the resistance, R, which is the ratio of the pressure difference produced by the pump to the rate of flow. A fatter pipe will have a lower resistance. Two pipes can be used in parallel, for instance when you turn on the water both in the kitchen and in the bathroom, and in this situation, the two pipes let more water flow than either would have let flow by itself, which tells us that they act like a single pipe with some lower resistance. The equation for their combined resistance is $R = 1/(1/R_1 + 1/R_2)$.
(a) Analyze the case where one resistance is fixed at some finite value, while the other approaches infinity. Give a physical interpretation.
(b) Likewise, discuss the case where one is finite, but the other becomes very small.

▷ Solution, p. 232

c1 Sketch the graph of the function $e^{-1/x}$, and evaluate the following four limits:

$$\lim_{x\to 0^+} e^{-1/x}$$
$$\lim_{x\to 0^-} e^{-1/x}$$
$$\lim_{x\to +\infty} e^{-1/x}$$
$$\lim_{x\to -\infty} e^{-1/x}$$

▷ Solution, p. 232

c2 Compute the following limits.

(a)

$$\lim_{x \to -4} (x+3)^{1492}$$

(b)

$$\lim_{x \to -4} (x+3)^{1493}$$

(c)

$$\lim_{x \to -\infty} (x+3)^{1493}$$

(d)

$$\lim_{x \to \infty} (\sin x)^{1492}$$

√

c3 Compute the following limits.

(a)

$$\lim_{u \to \infty} \frac{u^2+3}{u^2+4}$$

(b)

$$\lim_{u \to \infty} \frac{u^5+3}{u^2+4}$$

(c)

$$\lim_{u \to \infty} \frac{u^2+1}{u^5+2}$$

(d)

$$\lim_{u \to \infty} \frac{(2u+1)^4}{(3u^2+1)^2}$$

√

c4 Do the following notations make sense?

$$\lim_{x \nearrow \infty}$$

$$\lim_{x \searrow \infty}$$

$$\lim_{x \nearrow -\infty}$$

$$\lim_{x \searrow -\infty}$$

c5 Give two examples of functions for which $\lim_{x \searrow 0} f(x)$ does not exist.

c6 Find a constant k such that the function

$$f(x) = \begin{cases} 3x+2 & \text{for } x < 2 \\ x^2+k & \text{for } x \geq 2. \end{cases}$$

is continuous. Hint: Compute the one-sided limits. √

c7 A function f is defined by

$$f(x) = \begin{cases} x^3 & \text{for } x < -1 \\ ax+b & \text{for } -1 \le x < 1 \\ x^2+2 & \text{for } x \ge 1. \end{cases}$$

where a and b are constants. The function f is continuous. What are a and b? Hint: Compute the one-sided limits. ✓

c8 Find a rule for determining the number of horizontal and vertical asymptotes possessed by the following function.

$$f(x) = \frac{1}{ax^2+bx+c}$$

▷ Solution, p. 233

c9 Find any horizontal and vertical asymptotes of the following function.

$$f(x) = \frac{x^7+1234567}{x^7+1}$$

▷ Solution, p. 234

c10 Let

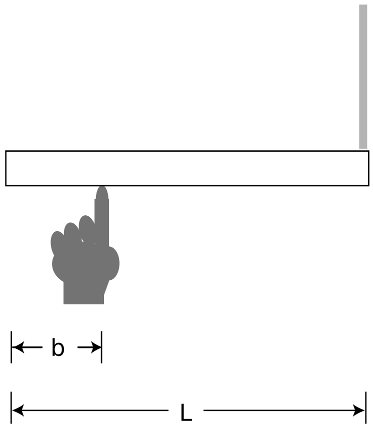
$$f(x) = \left(\frac{x^2+1}{x^2+2} - \frac{x^2+3}{x^2+4}\right)^{-1}.$$

Find any horizontal or vertical asymptotes. ▷ Solution, p. 234

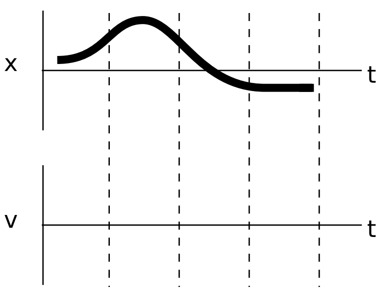
e1 The galactic empire has been pretty successful at crushing the rebel alliance, but there are still rebels laying low, scattered around in various solar systems. The empire offers a bounty x for the severed head of each rebel that is brought to the Dark Lord. Let f be the fraction of the rebels who are caught by the freelance bountry hunters. As in the examples in section 4.4.1, sketch the function $f(x)$ without knowing its equation. You should be able to infer whether or not $f'(0) = 0$. ▷ Solution, p. 234

e2 A pendulum is pulled back through an angle θ and then released. It then swings from θ to $-\theta$ and back to θ again; this is considered one complete oscillation. The time it takes to carry out this oscillation is called the period, T. If the pendulum is hung on a stiff rod rather than with a string, then θ can be as big as $180°$; you will find it helpful to consider what happens in the extreme case where θ *equals* $180°$. As in the examples in section 4.4.1, sketch the function $T(\theta)$ without knowing its equation. You should be able to infer whether or not $T'(0) = 0$.

e3 The rod in the figure is supported by the finger and the string. The tension T in the string depends on the distance b of the finger from the free end of the rod. As in the examples in section 4.4.1, sketch the function $T(b)$ without knowing its equation. The domain of the function consists of the physically possible values of b that allow the system to be in equilibrium. Discuss the x- and y-intercepts.



Problem e3.



Problem g1.

g1 The top part of the figure shows the position-versus-time graph for an object moving in one dimension. On the bottom part of the figure, sketch the corresponding velocity-versus-time graph.
▷ Solution, p. 235

i1 Let

$$f(x) = \frac{1}{x^2 - 4x + 5}$$

be defined on the interval $[-1, 1]$. Find any local and global extrema, as well as any asymptotes. Sketch the graph.

i2 Let

$$f(x) = \frac{1}{x^{10} - 1}.$$

Find any local and global extrema, as well as any asymptotes. Sketch the graph.

i3 Let

$$f(x) = \frac{x^2 + 1}{x - 1}.$$

Find any local and global extrema, as well as any asymptotes. Sketch the graph.

k1 Prove the following theorem. Let f be a real function whose second derivative is defined and continuous. If f'' is sometimes positive and sometimes negative, then f has a point of inflection x, and $f''(x) = 0$. Note that $f''(x) = 0$ is not the definition of a point of inflection, and that the theorem fails for a function on the rational numbers.
▷ Solution, p. 235

n1 Compute the following limits.
(a)

$$\lim_{t \to 1} \frac{t^2 + t - 2}{t^2 - 1}$$

(b)

$$\lim_{t \nearrow 1} \frac{t^2 + t - 2}{t^2 - 1}$$

(c)

$$\lim_{t \to -1} \frac{t^2 + t - 2}{t^2 - 1}$$

√

n2 Use the ϵ-δ definition to prove the following limit.

$$\lim_{x \to 3} x^2 = 9$$

Chapter 5
More derivatives

5.1 Transcendental numbers and functions

5.1.1 Transcendental numbers

Historically, the motivation for expanding the rational numbers to form the reals came from the desire to be able to discuss numbers like $\sqrt{2}$ or $\sqrt[3]{7}$. (The decision was not without controversy. Legend has it that Hippasus of Metapontum, who lived in the fifth century B.C., proved $\sqrt{2}$ to be irrational, and that the gods punished him by causing him to drown at sea.) We've already seen that the completeness property of the reals (section 4.5, p. 111) guarantees that $\sqrt{2}$ is a real number, and more generally one can use the intermediate value theorem to prove that roots of polynomials are real.

However, there are also numbers that cannot be defined as roots of polynomials having rational-number coefficients. These are called *transcendental* numbers. In some sense nearly all real numbers are transcendental. For example, suppose we generate a random digit by some method such as rolling dice, and we let this be the first digit in a decimal. Continuing in this way, we keep on generating more and more decimal places. If we could continue generating the digits indefinitely, then there would be a 100% probability that our number would be transcendental. The important mathematical constants π and e (the base of natural logarithms) are transcendental. Although transcendental numbers are the most common kind of real number, *proving* whether or not a particular number is transcendental can be difficult. Box 5.1 describes the first number that was ever proved to be transcendental. It was not until 44 years later that π was proved to be transcendental.

An important property of transcendental numbers is that they can't be written using any finite number of symbols in terms of rational numbers and the basic operations of arithmetic: addition, subtraction, multiplication, division, and roots. This is the reason for the name; transcendental numbers "transcend" arithmetic. For example, the number

$$\frac{-9+\sqrt{85}}{2} = 0.1098\ldots$$

is *not* transcendental, since it is written in terms of rational numbers and four of the basic operations. (It is a root of the polyno-

▷Box 5.1 **A transcendental number**

The first number proved to be transcendental, by Liouville in 1844, was:

$$0.110001000000000000000001\ldots$$

The first one occurs in the 1st decimal place, the next in the 2nd decimal place, the next in the 6th, and so on, with the sequence of numbers being 1, $1\cdot2 = 2$, $1\cdot2\cdot3 = 6$, ... Without going into the formal proof, it's not hard to get an intuitive feel for why this number is transcendental. Since the list of numbers 1, 2, 6, ... grows extremely rapidly, we find that as we continue to write the decimal expansion, it gets extremely sparse. It's so sparse that if we try to cook up a polynomial such as $P(x) = x^2 + 9x - 1$ with Liouville's number x as a root, we are bound to fail; x^2 and all higher powers of x are also extremely sparse, and this makes it impossible to get them to cancel out and give $P(x) = 0$. For a proof, see the Wikipedia article "Liouville number."

mial P given in box 5.1.) The converse is not true: not all non-transcendental numbers can be written using these operations. For example, the polynomial $x^5 - x + 1$ has a root $x \approx -1.17$, which cannot be expressed in terms of arithmetic.

5.1.2 Transcendental functions

Similarly, we have *functions* that are transcendental or not transcendental. For example, the function

$$f(x) = \frac{-9 + \sqrt{x}}{2}$$

is not transcendental because it can be written using the same basic operations of arithmetic. The techniques developed in chapter 2 are sufficient to differentiate any function that is not transcendental. The purpose of the present chapter is to see how to differentiate some functions that *are* transcendental.

Since the numbers π and e are transcendental, it is not surprising that the following closely related functions are transcendental:

$$\begin{aligned} &\sin x \\ &\cos x \\ &e^x \\ &\ln x \end{aligned}$$

Although the distinction between transcendental and non-transcendental *numbers* is of little practical significance (e.g., no real-world measurement will tell us whether a stick's length is transcendental or not), the distinction becomes an important one when we come to functions, because the methods we know so far will not suffice to differentiate a transcendental function. Most of this chapter will be concerned with how to extend our methods of differentiation to cover these functions.

5.2 Derivatives of exponentials

In example 3 on p. 19 and example 6 on p. 51 we found that the derivative of an exponential is an exponential: the more bunnies you have, the faster you produce baby bunnies; the more credit-card debt you have, the faster your debt grows. Furthermore, we were led to the conjecture that in the case of "the" exponential function e^x, the constant of proportionality between the function and its derivative was simply one:

$$(e^x)' = e^x \qquad (2)$$

There is no way to prove this unless we adopt some *definition* of e. In fact equation (2) serves as a perfectly good definition of e. Box 5.2 connects this to another popular definition.

▷Box 5.2 **A different definition of e**

Some people like lagers better than ales, Chicago better than Paris, and the following better than equation (2) as a definition of e:

$$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n \qquad (1)$$

The story-line behind (1) is something like this. Suppose your bank account carries an interest rate of 100%; the second 1 in the equation is 100/100. If the interest is compounded yearly, then your balance goes up every year by a factor of $(1 + 1/1)^1 = 2$. If it's compounded monthly at an interest rate of 100%/12, then the yearly increase is a factor of $(1+1/12)^{12} = 2.6$. If we let the 12 become a variable n that approaches infinity, then the 2.6 becomes e.

Let's connect this to equation (2). Applying the approximation $\mathrm{d}y/\mathrm{d}x \approx \Delta y/\Delta x$ to $y = e^x$, we have

$$e^x \approx 1 + x$$

for small values of x. Let $x = 1/n$, where n is large. Then $e^{1/n} \approx 1 + 1/n$, so $e \approx (1 + 1/n)^n$, which is consistent with equation (1).

Adopting equation (2) as a definition, application of the identity $b^x = e^{(\ln b)x}$ (see equation (9), p. 134) and the chain rule gives the more general rule

$$(b^x)' = (\ln b)b^x \tag{3}$$

for any base b.

Caffeine Example 1

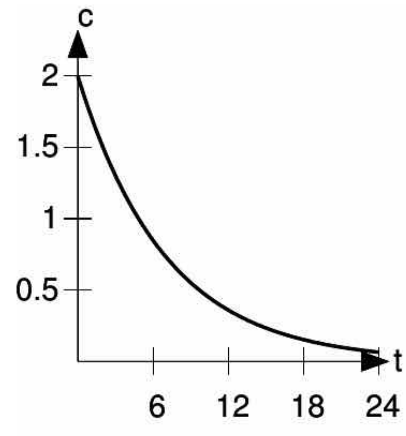
▷ The concentration of a foreign substance in the bloodstream generally falls off exponentially with time as $c = c_o e^{-t/a}$, where c_o is the initial concentration, and a is a constant. For caffeine in adults, a is typically about 7 hours. An example is shown in figure a. Differentiate the concentration with respect to time, and interpret the result. Check that the units of the result make sense.

▷ Using the chain rule,

$$\begin{aligned}\frac{dc}{dt} &= c_o e^{-t/a} \cdot \left(-\frac{1}{a}\right) \\ &= -\frac{c_o}{a} e^{-t/a}\end{aligned}$$

This can be interpreted as the rate at which caffeine is being removed from the blood and broken down by the liver. It's negative because the concentration is decreasing. According to the original expression for x, a substance with a large a will take a long time to reduce its concentration, since t/a won't be very big unless we have large t on top to compensate for the large a on the bottom. In other words, larger values of a represent substances that the body has a harder time getting rid of efficiently. The derivative has a on the bottom, and the interpretation of this is that for a drug that is hard to eliminate, the rate at which it is removed from the blood is low.

It makes sense that a has units of time, because the exponential function has to have a unitless argument, so the units of t/a have to cancel out. The units of the result come from the factor of c_o/a, and it makes sense that the units are concentration divided by time, because the result represents the rate at which the concentration is changing.



a / A typical graph of the concentration of caffeine in the blood, in units of milligrams per liter, as a function of time, in hours.

A base-10 exponential Example 2

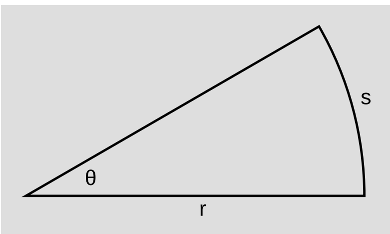
▷ Find the derivative of the function $y = 10^x$, verifying equation (3) directly in the case $b = 10$.

▷ In general, one of the tricks to doing calculus is to rewrite functions in forms that you know how to handle. This one can be rewritten as a base-e exponent:

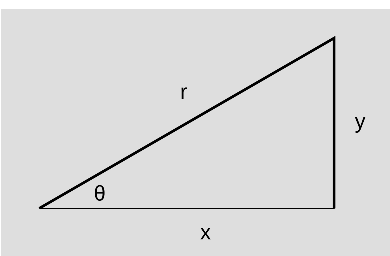
$$\begin{aligned} y &= 10^x \\ \ln y &= \ln\left(10^x\right) \\ \ln y &= x \ln 10 \\ y &= e^{x \ln 10} \end{aligned}$$

Applying the chain rule, we have the derivative of the exponential, which is just the same exponential, multiplied by the derivative of the inside stuff:

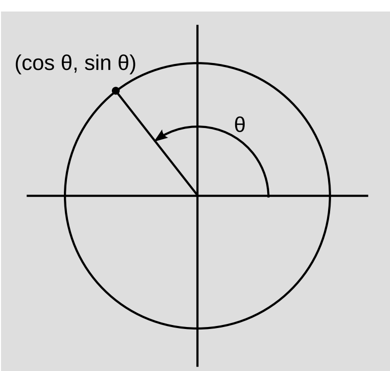
$$\begin{aligned} \frac{dy}{dx} &= e^{x \ln 10} \cdot \ln 10 \\ &= (\ln 10) 10^x \end{aligned}$$



b / The radian measure of the angle θ is s/r.



c / The sine of θ is y/r, the cosine x/r.



d / The sine and cosine defined on the unit circle, for any angle θ.

5.3 Review: the trigonometric functions

Before we talk about how to differentiate trig functions, here's an opportunity to refresh your memory on what trig functions are in the first place.

5.3.1 Radian measure

The presence of numbers like 60 and 360 in our units of measurement for time and angles dates back to the ancient Babylonians. The reason for splitting larger quantities into these numbers of subdivisions is that 60 and 360 are divisible by many small integers, including 2, 3, 5, 10, and 12. For practical purposes it's fine for a carpenter to define a right angle as $90\,°$. But it turns out to be much less cumbersome when doing calculus to adopt the radian as our unit of angle, as defined in figure b. A right angle is $\pi/2$ radians, a full circle 2π. From the definition we observe that a number with "units" of radians is in fact the unitless ratio of two distances.

5.3.2 Sine and cosine

Figure c shows a right triangle. The sine and cosine of the angle θ are defined as the ratios

$$\sin\theta = \frac{y}{r} \qquad \text{and}$$
$$\cos\theta = \frac{x}{r}.$$

Since these ratios are the same for any two similar triangles, the definitions depend only on θ, not on the triangle.

5.3.3 Arbitrary angles

Since the above definition assumes a right triangle, it is restricted to angles θ that are between 0 and $\pi/2$ (a right angle). Figure d shows how to generalize this to an angle that is an arbitrary real number. The circle is the *unit* circle, i.e., the circle centered on the origin and having radius 1. The angle is by convention measured counterclockwise from the x axis; a negative angle would indicate a clockwise rotation. The (x, y) coordinates of a point on the unit circle at angle θ are $(\cos\theta, \sin\theta)$.

It is handy to know these facts:

$$\cos 0 = 1$$
$$\sin 0 = 0$$

These do not need to be memorized. They can be recovered instantly by visualizing the unit circle.

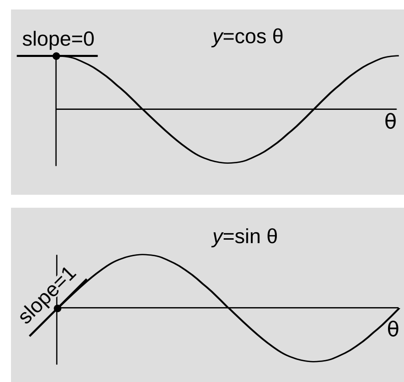
The following identities will be needed later in the chapter.

$$\sin(x + y) = \sin x \cos y + \cos x \sin y \tag{4a}$$
$$\cos(x + y) = \cos x \cos y - \sin x \sin y \tag{4b}$$

5.3.4 Other trigonometric functions

In terms of the same variables defined above, we have the following additional trigonometric functions:

$$\tan\theta = \frac{y}{x} \qquad \text{[important]}$$
$$\csc\theta = 1/\sin\theta \qquad \text{[not as important]}$$
$$\sec\theta = 1/\cos\theta \qquad \text{[not as important]}$$
$$\cot\theta = 1/\tan\theta \qquad \text{[not as important]}$$



e / The derivatives of the cosine and sine functions at $\theta = 0$.

5.4 Derivatives of trigonometric functions

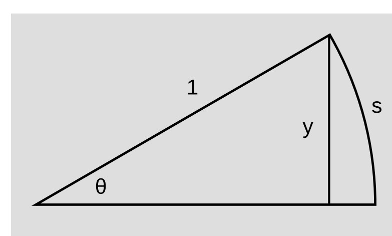
5.4.1 Derivatives of the sine and cosine

Sometimes a variable oscillates back and forth. A weight hung from a rubber band will vibrate up and down. The temperature of Los Angeles goes down every winter and back up every summer. A sinusoidal wave is the most mathematically simple model of such an oscillation, and if we want to know the rate of change, we need to know how to differentiate such a function.

So how would we find the derivative of a sine or cosine? Since they're transcendental, they can't be expressed in terms of simpler functions that we know how to differentiate.

Derivatives at $\theta = 0$

Let's start by finding the derivatives of these functions at zero, as shown in figure e.



f / A geometrical method of finding $\sin' 0$.

Since the cosine is an even function, we have $\cos' 0 = 0$.

What about $\sin' 0$? The definition of the derivative gives

$$\begin{aligned}\sin' 0 &= \lim_{\theta\to 0} \frac{\sin\theta - \sin 0}{\theta - 0} \\ &= \lim_{\theta\to 0} \frac{\sin\theta}{\theta}.\end{aligned}$$

In figure f, the definition of radian measure gives $\theta = s$, while the definition of the sine function tells us that $\sin\theta = y$. Thus the limit above becomes

$$\sin' 0 = \lim_{\theta\to 0} \frac{y}{s}.$$

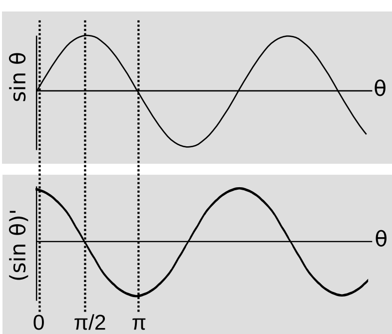
If θ is close to zero, then the lengths y of the vertical line and s of the arc should be nearly the same, so we have the small-angle approximation $\sin\theta \approx \theta$. Our limit is clearly[1] equal to 1, so we have $\sin' 0 = 1$.

As a check on our work, we can take a numerical approximation to the derivative at $\theta = 0$,

$$\begin{aligned}\sin' 0 &\approx \frac{\sin 0.001 - \sin 0}{0.001} \qquad \text{[angle in radians]} \\ &= 0.99999983,\end{aligned}$$

which is indeed close to 1.

A preliminary sketch



g / Sketching the derivative of the sine function.

What about the value of $\sin'$ at $\theta \neq 0$? Let's *sketch* the derivative of $\sin\theta$ in order to gain some insight. Using the techniques of section 4.4.2, p. 108, we obtain figure g. At $\theta = 0$, the slope of the sine function is 1, which is as large and positive as it ever gets, so the value of the derivative sketched in the bottom graph is large and positive. At $\pi/2$ (90 degrees), the sine has its maximum value of 1, and its derivative is 0. At π, the sine has its largest negative derivative. The graph we're led to draw for $\sin'\theta$ looks like the cosine function.

The graph of the cosine function is the same as the graph of the sine function except for a shift to the left by a quarter of a cycle. Therefore by the shift property of the derivative (p. 16), if the derivative of sin is cos, then the derivative of cos must be a cosine function shifted to the left by another quarter-cycle, which gives $-\sin$. Curve sketching therefore leads us to the following conjectures:

$$\begin{aligned}\sin' &= \cos \\ \cos' &= -\sin\end{aligned}$$

[1] Strictly speaking, we should prove that for the approximation $\sin\theta \approx \theta$, the error $E = \theta - \sin\theta$ goes to zero fast enough so that $\lim_{\theta\to 0} E/\theta = 0$. In fact, one can show based on the *areas* in figure f that $|E| < |\theta^2|$ for $|\theta| < 0.1$.

Proof of the derivatives of the sine and cosine

To prove this, let's apply the definition of the derivative to the sine function.

$$\sin' x = \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h}$$

Making use of the identity $\sin(x+y) = \sin x \cos y + \cos x \sin y$ (p. 129), we find

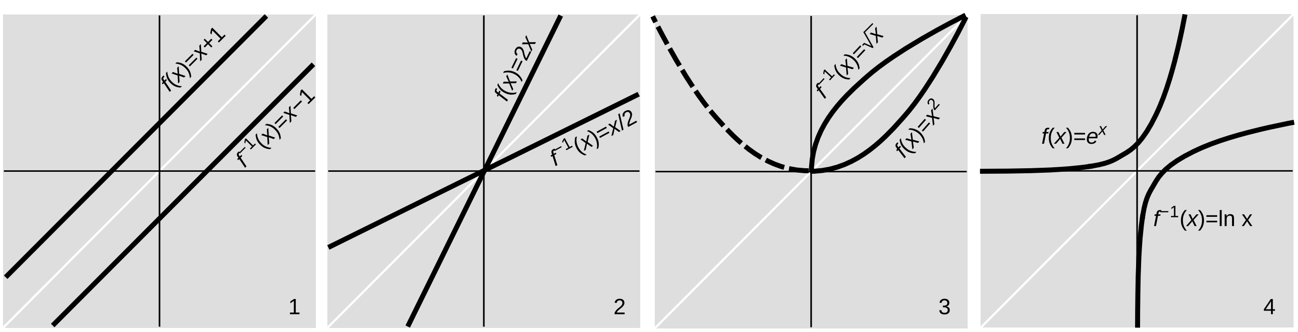
$$\begin{aligned} \sin' x &= \lim_{h\to 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\ &= \cos x \lim_{h\to 0} \frac{\sin h}{h} + \sin x \lim_{h\to 0} \frac{\cos h - 1}{h}. \end{aligned}$$

We have already determined these two limits: they are 1 and 0, respectively, so $\sin' x = \cos x$ as claimed. The similar calculation for the derivative of $\cos x$ is left as an exercise.

5.5 Review: the inverse of a function

Some operations can be undone. Others can't. Computer software often has an "undo" function. But what if the operation is mixing hot coffee with cold milk? There is no way to undo this operation, even in principle, because information has been lost. No matter how closely we inspect the mixture, we have no way of determining how hot the original coffee was, or how cold the original milk.

We've defined a function as a graph that passes the vertical line test, so that every input x corresponds to a single output y. A function may or may not be undoable. If every y corresponds to a single x, i.e., if the function passes a *horizontal* line test, then it's undoable, and we call the "undo" operation the inverse of the function. The inverse of a function f is notated f^{-1}, where only context tells us that we mean the "undoing" of f, rather than $1/f$.



h / Some functions and their inverses. In each case, the inverse function is found by reflecting the graph across the line $y = x$.

Geometrically, inverting the function means interchanging the roles of x and y, which requires flipping it across the 45-degree diagonal defined by the line $y = x$, as in figure h. For example, figure h/1 shows the "add-one" function defined by $f(x) = x + 1$, and the "subtract-one" function $f^{-1}(x) = x - 1$ that undoes it.

We define a function as a graph that passes the vertical-line test. The set of all x values for which the graph contains an (x, y) point is called the *domain* of the function, while the set of such y values is its *range*. That is, the domain is the set of all legal inputs, while the range is the set of possible outputs. Sometimes we define a particular function using a formula, and this may implicitly restrict its domain. For example, if we define

$$y = \frac{1}{x-1},$$

then by implication the domain is the whole real line except for $x = 1$, which would produce division by zero.

Sometimes there are real-world reasons for restricting the domain of a function. For example, in section 4.3.3, p. 103, we discussed the amount of tension T in a telephone wire that was necessary in order to make it sag by a height h at the middle. This function was of the form $T = k/h$, where k is a constant. Mathematically this function is well defined for $h < 0$, but physically that would be meaningless, since a cable can only sustain tension ($T > 0$) — only a rigid object such as a rod can sustain compression ($T < 0$).

Sometimes by restricting the domain of a function we can make it invertible. For example, the function $y = x^2$ fails the horizontal-line test, so it doesn't have an inverse function. But if we restrict its domain to $x \geq 0$, as in figure h/4, then we can define its inverse function, which is $x = \sqrt{y}$ (using the positive root).

In terms of the composition of functions (section 2.4.3, p. 56), the function $f \circ f^{-1}$ is simply the identity function $y = x$ (perhaps with a restriction on its domain and range). The same applies to $f^{-1} \circ f$.

Discussion question

A Which of the following four statements are true, and which are false?

1. For all real numbers x, $\sin(\sin^{-1} x) = x$.
2. For all real numbers x, $\sin^{-1}(\sin x) = x$.
3. For all real numbers x, $\tan(\tan^{-1} x) = x$.
4. For all real numbers x, $\tan^{-1}(\tan x) = x$.

5.6 Derivative of the inverse of a function

Suppose that x is how many gallons of gas I buy, and y is how much money I pay. Then y is a function of x, and the rate at which this function changes, i.e., the price per gallon of gas, in my area is currently about

$$\frac{\Delta y}{\Delta x} = 4\ \frac{\$}{\text{gallon}}.$$

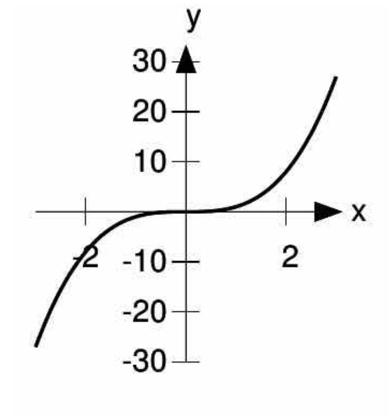
It's valid to measure this rate of change with an expression of the form $\Delta \ldots / \Delta \ldots$, because the rate of change is constant. I might also want to know how much gas I can get for each additional dollar I'm willing to spend, and this is found by ordinary algebra to be

$$\frac{\Delta x}{\Delta y} = 0.25\ \frac{\text{gallon}}{\$}.$$

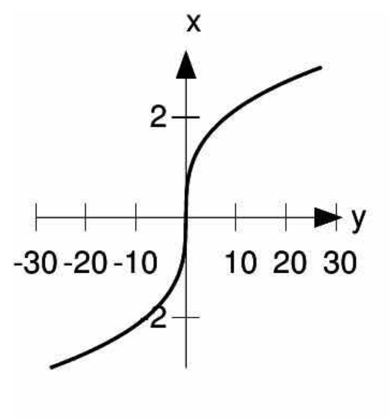
If y is a function of x, and the function is invertible, then the Leibniz notation suggests that this should hold even for non-constant rates of change, i.e., that the derivative of the inverse function is

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)}.$$

This is in fact correct, with the caveat that when $\mathrm{d}y/\,\mathrm{d}x = 0$, $\mathrm{d}x/\,\mathrm{d}y$ is undefined because it blows up to infinity.



i / The function $y = x^3$.



j / The function $x = y^{1/3}$.

Derivative of a cube root Example 3

▷ Let $y = x^3$. Find $\mathrm{d}x/\,\mathrm{d}y$.

▷ The function $y = x^3$, figure i, has a well-defined inverse $x = y^{1/3}$, which is the cube root, figure j. The derivative of the original function is

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 3x^2.$$

The derivative of the inverse function is

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}y} &= \frac{1}{\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)}\\ &= \frac{1}{3x^2}\\ &= \frac{1}{3}x^{-2}.\end{aligned}$$

If we prefer to express this in terms of y, we can substitute to get

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{3}y^{-2/3},$$

which agrees with the power rule (section 2.6, p. 57).

This expression holds everywhere except $x = 0$, $y = 0$, where $\mathrm{d}x/\,\mathrm{d}y$ blows up to infinity.

5.7 Review: logarithms

5.7.1 Logarithms

The inverse of exponentiation is the logarithm. If

$$b^p = z,$$

then

$$\log_b z = p.$$

For example, $\log_2 8 = 3$, because $2^3 = 8$.

The number 10 has appeared above as a base, and that's because humans have 10 fingers. There's clearly nothing all that special about 10. It's an accident of evolution. A number with more cosmic significance is $e \approx 2.71818\ldots$ Exponents and logarithms with base e have some nice properties, which we'll discuss later in more detail. Any expression with x in the exponent is called an exponential, but e^x is "the" exponential function. Sometimes when x is a complicated expression it gets awkward to write it as a superscript, and then we write $\exp(\ldots)$ instead of $e^{\cdots}$. The logarithm with the special base e is called the *natural logarithm*, notated ln.

5.7.2 Identities

The following identities are useful. Exponentials and logs are inverse operations:

$$\log_b (b^x) = x \tag{5a}$$

$$b^{\log_b x} = x \tag{5b}$$

Logs turn multiplication and division into addition and subtraction:

$$\log(xy) = \log x + \log y \tag{6a}$$

$$\log(x/y) = \log x - \log y \tag{6b}$$

A log in one base can be changed into a log in another base:

$$\log_b x = \frac{\log_c x}{\log_c b} \tag{7}$$

For example, $\log_{10} 10^6 = 6$, whereas $\log_{100} 10^6 = 3$. It may be convenient to convert a logarithm to a natural log, with $c = e$:

$$\log_b x = \frac{\ln x}{\ln b}. \tag{8}$$

Similarly, an exponential with an arbitrary base b can be converted to an exponential with base e.

$$b^x = e^{(\ln b)x} \tag{9}$$

5.8 The derivative of a logarithm

We now know enough to differentiate a logarithm. The natural log has the nicest properties, so we'll start with it. Let

$$y = \ln x.$$

Then

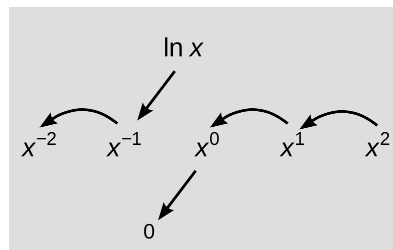
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{1}{\left(\frac{\mathrm{d}x}{\mathrm{d}y}\right)} && \text{[derivative of an inverse]}\\
&= \frac{1}{\left(\frac{\mathrm{d}e^y}{\mathrm{d}y}\right)} && [x = e^y]\\
&= \frac{1}{e^y} && \text{[derivative of the exponential is the exponential]}\\
&= \frac{1}{x} && [x = e^y \text{ again}]
\end{aligned}$$

The result is unexpectedly simple.

Derivative of the natural logarithm

$$\frac{\mathrm{d}\,\ln x}{\mathrm{d}x} = \frac{1}{x}$$

This is noteworthy because it shows that there must be an exception to the rule that we can always obtain a function that varies like x^{n-1} by differentiating something like x^n. If we believed that this rule was always true, then we would think that we could obtain the function x^{-1} by differentiating some function of the form (constant)x^0. But in fact this doesn't work, since x^0 is a constant, and the derivative of x^0 is therefore 0. Figure k shows the idea.



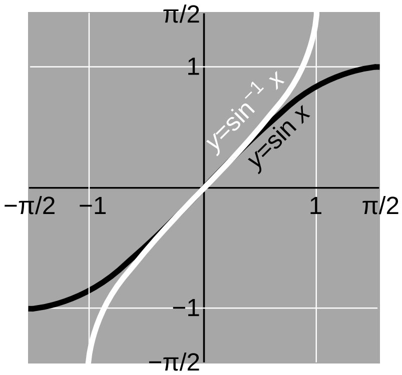
k / A "ladder" of powers of x. Ignoring multiplicative constants, differentiation usually just takes us one step down the ladder. The diagram shows the two exceptions.

Derivatives of logs with other bases can be found by using equation (8) to convert to a natural log. The result is

$$\frac{\mathrm{d}\,\log_b x}{\mathrm{d}x} = \frac{1}{(\ln b)x}$$

The power rule for irrational exponents Example 4

In section 2.6, p. 57, we showed that the power rule $(x^n)' = nx^{n-1}$ held for any nonzero integer value of n, and also gave a sample of a proof for a fractional exponent. However, the methods used there were not capable of proving the result for irrational values of n, or of demonstrating it for all rational values in a single proof. We now have the ability to carry out the proof in an efficient way for any real, nonzero n.



l / The sine and inverse sine functions.

$$
\begin{aligned}
y &= x^n \\
&= e^{n\ln x}
\end{aligned}
$$

By the chain rule,

$$
\begin{aligned}
\frac{dy}{dx} &= e^{n\ln x}\cdot\frac{n}{x} \\
&= x^n\cdot\frac{n}{x} \\
&= nx^{n-1}.
\end{aligned}
$$

(For $n = 0$, the result is zero.)

5.9 Derivatives of inverse trigonometric functions

The sine and cosine functions are not invertible, since they fail the horizontal line test — in fact, any horizontal line that crosses these functions crosses them in infinitely many places. For example, if I tell you that I took the sine of some angle, and the sine was zero, then the angle could have been any number from the infinite set $\{\ldots-2\pi,-\pi,0,\pi,2\pi,\ldots\}$. But by restricting the domain of the sine function appropriately, e.g., to $-\pi/2\le x\le\pi/2$, we can make an invertible function and define an inverse sine, figure l.

The derivative of the inverse sine can be found straightforwardly by using our knowledge of the derivatives of inverses of functions. Let $y=\sin^{-1}x$. Then:

$$
\begin{aligned}
\frac{dy}{dx} &= \frac{1}{\left(\frac{dx}{dy}\right)} \\
&= \frac{1}{\cos y} \qquad \text{[because } x=\sin y\text{]} \\
&= \frac{1}{\sqrt{1-\sin^2 y}} \qquad \text{[because } (\cos y,\sin y) \text{ lies on the unit circle]} \\
&= \frac{1}{\sqrt{1-x^2}}
\end{aligned}
$$

A similar calculation shows that the derivative of $\cos^{-1}x$ is $-1/\sqrt{1-x^2}$.

5.10 Summary of derivatives of transcendental functions

Given the derivatives of trig and inverse trig functions from sections 5.4 and 5.9, it is straightforward to extend the list of derivatives to include the other familiar trig functions. In this section we provide a summary for reference purposes of all of the derivatives of the transcendental functions encountered so far.

$(e^x)' = e^x$	$(\ln x)' = 1/x$
$(\sin x)' = \cos x$	$(\sin^{-1} x)' = (1-x^2)^{-1/2}$
$(\cos x)' = -\sin x$	$(\cos^{-1} x)' = -(1-x^2)^{-1/2}$
$(\tan x)' = (\cos x)^{-2}$	$(\tan^{-1} x)' = (1+x^2)^{-1}$

5.11 Hyperbolic functions

The hyperbolic trig functions are defined as follows.

$$\begin{aligned} \sinh x &= \frac{1}{2}\left(e^x - e^{-x}\right) \\ \cosh x &= \frac{1}{2}\left(e^x + e^{-x}\right) \qquad \text{and} \\ \tanh x &= \frac{\sinh x}{\cosh x}. \end{aligned}$$

Their inverses can be calculated using the following relations:

$$\begin{aligned} \sinh^{-1} x &= \ln\left(x + \sqrt{x^2+1}\right) \\ \cosh^{-1} x &= \ln\left(x + \sqrt{x^2-1}\right) \\ \tanh^{-1} x &= \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) \end{aligned}$$

The derivatives are as follows:

$(\sinh x)' = \cosh x$	$(\sinh^{-1} x)' = (x^2+1)^{-1/2}$
$(\cosh x)' = \sinh x$	$(\cosh^{-1} x)' = (x^2-1)^{-1/2}$
$(\tanh x)' = (\cosh x)^{-2}$	$(\tanh^{-1} x)' = (1-x^2)^{-1}$

Review problems

a1 For what set of angles θ do we have both $\sin\theta < 0$ and $\cos\theta < 0$? ▷ Solution, p. 235

a2 Let the function f be defined by $f(x) = x^3 + 1$. Find an expression for the function f^{-1}. ✓

a3 Evaluate $\log_3 \sqrt{1/27}$. ✓

Problem b1 does not require any of the new calculus learned in this chapter, but does require knowledge of the transcendental functions reviewed in it.

b1 Find the following limits at infinity. Check your results by plugging in large numbers on a calculator or by graphing.

(a)

$$\lim_{x\to\infty} \frac{\sin x}{\sin(x+\pi)}$$

(b)

$$\lim_{x\to\infty} \frac{\sqrt{x+1}\,\cos x}{x+3}$$

(c)

$$\lim_{x\to\infty} \frac{\ln x}{x}$$

(d)

$$\lim_{x\to\infty} \frac{e^{-x}}{\cos x}$$

✓

Problems

c1 Differentiate $\ln(2t+1)$ with respect to t.
▷ Solution, p. 235

c2 Differentiate $a \sin(bx+c)$ with respect to x.
▷ Solution, p. 235

c3 Differentiate the following with respect to x: e^{7x}, e^{e^x}. (In the latter expression, as in all exponentials nested inside exponentials, the evaluation proceeds from the top down, i.e., $e^{(e^x)}$, not $(e^e)^x$.)
▷ Solution, p. 235

c4 The range of a gun, when elevated to an angle θ, is given by

$$R = \frac{2v^2}{g} \sin\theta \, \cos\theta .$$

Find the angle that will produce the maximum range.
▷ Solution, p. 236

c5 Prove, as claimed on p. 137, that the derivative of $\tan\theta$ with respect to θ is $(\cos\theta)^{-2}$. Assume that the derivatives of the sine and cosine are already known.
▷ Solution, p. 236

c6 Show that the function $\sin(\sin(\sin x))$ has maxima and minima at all the same places where $\sin x$ does, and at no other places.
▷ Solution, p. 236

c7 Find any extrema of the hyperbolic cosine function defined on p. 137.
▷ Solution, p. 237

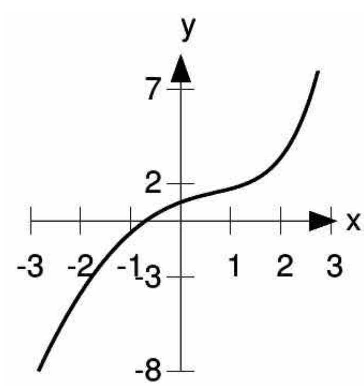
d1 (a) Let $y = \ln(1+x)$. Find the best linear approximation to this function near $x = 0$. ✓
(b) Use the result of part a to approximate the value of $\ln(1.003)$ without a calculator. ✓

d2 (a) Let $y = \cos x$. Find the best linear approximation to this function near $x = \pi/2$. ✓
(b) Use the result of part a to approximate the value of $\cos(1.5)$ without a calculator. ✓

d3 (a) Use the graph to visually estimate the location of the inflection point of the function

$$y = e^x - x^2 .$$

(b) Use calculus to find the point exactly. ✓



Problem d3.

d4 The function

$$y = 3^x - 2^{-x}$$

has one inflection point. Locate it. $\checkmark$

In problems e1-e4, differentiate the given functions.

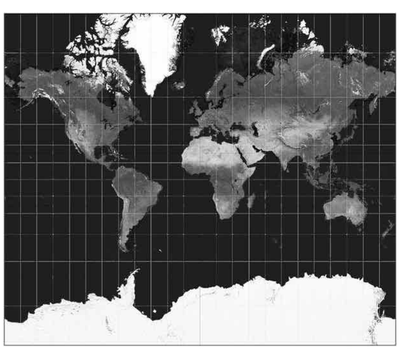
e1 $\sin \cos \tan x$ $\checkmark$

e2 $\ln \cos e^x$ $\checkmark$

e3 $\exp \sin \ln x$ $\checkmark$

e4 $\tan^{-1} \sqrt{\ln x}$ $\checkmark$

e5 Differentiate the function x^x. $\checkmark$



A Mercator projection, problem e6. Note the extremely exaggerated scale at the poles.

e6 On a map drawn using a Mercator projection, the y coordinate on the paper is given by $y = a \tanh^{-1} \sin \phi$, where ϕ is the latitude, a is a constant, and the inverse hyperbolic tangent function is defined on p. 137. (a) Find the derivative $\mathrm{d}y/\,\mathrm{d}\phi$, which indicates the latitude-dependent scale of the map in the north-south direction. (b) The approximations $\tanh x \approx x$ and $\sin x \approx x$ are valid for small x. Use these approximations to approximate the behavior of $y(\phi)$ for small ϕ, and use this to check your answer to part a. $\checkmark$

f1 A cold bottle of beer is left outside under a shady tree at a picnic. Its temperature as a function of time is given by

$$T = a - be^{-ct},$$

where a, b, and c are constants.
(a) Infer the units of a, b, and c. (For examples of how to do this, see section 1.9 on p. 34, example 9 on p. 29, and example 1 on p. 127.)
(b) Find the derivative $\mathrm{d}T/\,\mathrm{d}t$, which measures how fast the beer is warming up. Check that its units make sense.
(c) Interpret both the original equation and your answer to part b in the limit where $t \to \infty$.
(d) Interpret the constants a, b, and c physically.

▷ Solution, p. 237

f2 A person is parachute jumping. During the time between when she leaps out of the plane and when she opens her chute, her altitude is given by an equation of the form

$$y = b - c\left(t + ke^{-t/k}\right).$$

where b, c, and k are constants. Because of air resistance, her velocity does not increase at a steady rate as it would for an object falling in vacuum.
(a) What units would b, c, and k have to have for the equation to make sense? (For examples of how to do this, see section 1.9 on p. 34, example 9 on p. 29, example 1 on p. 127, and problem f1 above.)
(b) Find the person's velocity, v, as a function of time. ✓
(c) Use your answer from part b to get an interpretation of the constant c.
(d) Find the person's acceleration, a, as a function of time. ✓
(e) Use your answer from part d to show that if she waits long enough to open her chute, her acceleration will become very small.

f3 If an object is vibrating, and the vibration is gradually dying out, its motion (position as a function of time) is typically of the form

$$x(t) = A\cos(\omega t + \delta)e^{-bt},$$

where A, ω, δ, and b are constants.
(a) Infer the units of each of the four constants, and give a physical interpretation. (For examples of how to infer the units, see section 1.9 on p. 34, example 9 on p. 29, example 1 on p. 127, and problem f1 above.)
(b) Find the velocity.
(c) Check that the units of your answer to part b make sense. ✓

f4 Sometimes doors are built with mechanisms that automatically close them after they have been opened. The designer can set both the strength of the spring and the amount of friction. If there is too much friction in relation to the strength of the spring, the door takes too long to close, but if there is too little, the door will oscillate. For an optimal design, we get motion of the form

$$x = cte^{-bt},$$

where x is the position of some point on the door, and c and b are positive constants. (Similar systems are used for other mechanical devices, such as stereo speakers and the recoil mechanisms of guns.) In this example, the door moves in the positive direction up until a certain time, then stops and settles back in the negative direction, eventually approaching $x = 0$. This would be the type of motion we would get if someone flung a door open and the door closer then brought it back closed again. (a) Infer the units of the constants b and c. (For examples of how to do this, see example 9 on p. 29, example 1 on p. 127, and problem f1 above.)
(b) Find the door's maximum speed (i.e., the greatest absolute value of its velocity) as it comes back to the closed position. ✓
(c) Show that your answer has units that make sense.

g1 Credit card fraud creates costs (including both economic costs and inconvenience) for businesses, credit card holders, and the credit card companies. If the company institutes a particular measure to prevent fraud, it may be able to eliminate some fraction of the fraud that would otherwise have occurred. Putting some additional measure in place may then eliminate some fraction of the remaining fraud, further reducing the total amount. Let the amount the company spends on prevention be p. For the reasons described above, it's reasonable to imagine that fraud falls off exponentially as a function of p, so that the total cost to the company is

$$C(p) = p + ae^{-bp}.$$

Here a and b are constants, the first term represents the cost of carrying out the fraud prevention, and the second term represents the cost of the fraud that was not prevented.
(a) Find the value of p that minimizes the cost. ✓
(b) Check that the units of your answer make sense (section 1.9, p. 34).
(c) For what values of the parameters a and b does your answer not produce a meaningful result? Check that this makes sense.
(d) Suppose that legislation forces the credit card company to suffer more of the consequences of the fraud, rather than making their customers bear the brunt. What change does this imply in the parameters of the model? Check that your answer to part a shows the right trend when this change is applied.

g2 Benjamin Gompertz (1779-1865) was a British mathematician and pioneering actuarial scientist, who overcame significant social barriers due to antisemitism. We would all like to live forever, and actuaries are in the business of telling us that we probably can't. Based on mortality data, Gompertz constructed a model in which an initial population N_o of babies born at $t = 0$ becomes at a later time t a surviving population

$$N = N_o e^{1-e^t},$$

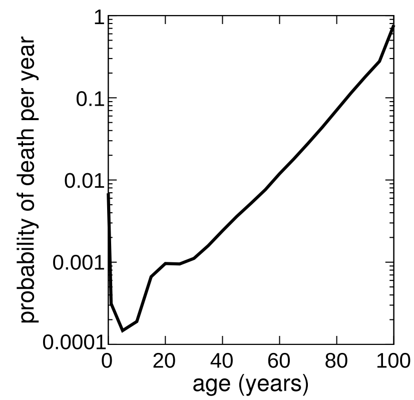
where I've simplified the expression by leaving out some constants. If you've survived to age t, then your probability of dying in the coming year is

$$-\frac{\Delta N}{N},$$

where $-\Delta N$ is the number of deaths per year. Therefore the death rate is

$$-\frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}t}.$$

Show that in the Gompertz model, this death rate is proportional to e^t. This exponential rate of increase is demonstrated in the figure.



Problem g2. Probability of death in the U.S. in the year 2003. Note the logarithmic scale on the vertical axis. Between the ages of about 30 and 95, the death rate rises exponentially, as shown by the linearity of the data on the logarithmic graph.

g3 In problem g1 on p. 142, we minimized a function that looked like

$$y = x + ae^{-bx},$$

where x, a, and b were all positive. Suppose instead that the function had been

$$y = x^2 + ae^{-bx},$$

with the corresponding quantities still being positive. Using the same technique to find its minimum, we obtain an equation of a type called a transcendental equation, which cannot be solved exactly for x in terms of elementary functions. Use the intermediate value theorem to prove that such a minimum nevertheless exists, as long as a and b are both greater than zero.

k1 Proof by induction was introduced in section 2.6.1, p. 58. Use induction to prove that

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n} b^x = (\ln b)^n b^x.$$

To understand what's going on, you may wish to calculate the first few derivatives; however, doing this and observing the pattern does not constitute a proof.

k2 The function

$$f(x) = e^{-\frac{1}{2}x^2}$$

defines the standard "bell curve" of statistics. (Note that exponentiation is not associative, and that in exponentiation, x^{y^z} means $x^{(y^z)}$, not $(x^y)^z$; an expression of the latter form is not very interesting, since it simply equals $x^{(yz)}$.)

Proof by induction was introduced in section 2.6.1, p. 58. Use induction to prove that the nth derivative of f is of the form

$$f^{(n)}(x) = P_n(x)e^{-\frac{1}{2}x^2},$$

where P_n is an nth order polynomial. To understand what's going on, you may wish to calculate the first few derivatives; however, doing this and observing the pattern does not constitute a proof.

Chapter 6

Indeterminate forms and L'Hôpital's rule

6.1 Indeterminate forms

6.1.1 Why 1/0 and 0/0 are not morally equivalent

If you enter $1/0$ and $0/0$ into your calculator, it probably flashes the same error message in both cases. You learned in grade school that division by zero is "undefined." But there are completely different reasons why these two types of division by zero are undefined. Briefly:

- $1/0$ is undefined as a real number because it would have to be infinite, and the real number system doesn't include infinite numbers.[1]
- $0/0$ is undefined because writing this expression doesn't give enough information to say what it equals.

Suppose that for some real number x, we had

$$\frac{0}{0} = x.$$

Multiplying by 0 on both sides gives a condition

$$0 = 0x$$

that x should satisfy. But *every* real number has this property, so writing $0/0$ doesn't give enough information to say whether x is defined and, if so, what its value is. Expressions of this "not-enough-information" type are called *indeterminate forms.*

6.1.2 Indeterminate forms from brute force on a limit

When we try to evaluate a limit, usually our first attempt is simply to plug in and see if a number comes out. For example, if we want to evaluate

$$\lim_{x \to 0} \frac{1+x}{3+x},$$

we will naturally try plugging in $x = 0$, get the result $1/3$, and we're done. This is not an indeterminate form. But, for example, suppose

[1] See section 2.9, p. 64, and example 11, p. 113.

▷Box 6.1 **More indeterminate forms**

We will mainly be concerned with the indeterminate form $0/0$, but there are other ones as well. Suppose we try to evaluate the limit

$$\lim_{\theta \nearrow \pi/2} \left(\frac{\pi}{2} - \theta\right) \tan\theta$$

by plugging in $\theta = \pi/2$. This fails because the first factor goes to zero, but the tangent factor blows up to infinity. This is an example of the indeterminate form $0 \cdot \infty$. The limit is defined and equals 1, but plugging in won't tell us that.

The limit

$$\lim_{x \to \infty} \sqrt{x+1} - \sqrt{x-1}$$

is an example of the indeterminate form $\infty - \infty$. It equals zero.

that $f(x) = x^2$ and we want to evaluate $f'(1)$. The definition of the derivative in terms of a limit gives

$$\lim_{h \to 0} \frac{(1+h)^2 - 1}{h},$$

and attempting to plug in $h = 0$ results in the indeterminate form $0/0$. This limit is well defined; it equals 2. But the indeterminate form tells us that the brute-force technique was too crude, and we needed to handle the calculation a little more delicately.

The indeterminate form $0/0$ can also be undefined. For example, $\lim_{x \searrow 0} \frac{\sqrt{x}}{x^2} = \infty$.

6.2 L'Hôpital's rule in its simplest form

Every derivative, if defined, can be seen as a case of the indeterminate form $0/0$. Conversely, we can often convert a $0/0$-type limit into a problem in evaluating derivatives. Suppose that we want to calculate a limit of the form

$$\lim_{x \to a} \frac{u(x)}{v(x)},$$

where $u(a) = 0$ and $v(a) = 0$. Then $\Delta u = u(x) - u(a)$ means the same thing as u, and similarly, Δv equals v. So we can rewrite our limit as

$$\lim_{x \to a} \frac{\Delta u}{\Delta v},$$

or

$$\lim_{x \to a} \frac{\Delta u / \Delta x}{\Delta v / \Delta x}.$$

If $v'(a) \neq 0$, then by property P_6 of the limit, p. 95, our limit becomes

$$\frac{\lim_{x \to a} \Delta u / \Delta x}{\lim_{x \to a} \Delta v / \Delta x},$$

which equals

$$\frac{u'(a)}{v'(a)}.$$

We have proved the following.

a / Guillaume de L'Hôpital (1661-1704) was a French marquis. Born into a military family, he eventually became a mathematician because of bad eyesight. He wrote the first calculus textbook. As acknowledged in the preface, the results given in the book originated with Leibniz and the Bernoulli brothers, but L'Hôpital's own name has become attached to the theorem known as L'Hôpital's rule. When students meet the Marquis, they always wonder about his name, which looks like the English word "hospital." Actually, he spelled it with an "s," and it is the same word in French. The "H" is silent, and the accent is on the "a." As French people gradually stopped pronouncing the "s," they stopped writing it, but put the housetop accent on the "ô" to show what they were leaving out. The family name probably comes from an early association with a "hospital," a word that in medieval times had a broader meaning, encompassing institutions such as guest-houses for pilgrims and what we would today call subsidized public housing.

Theorem: L'Hôpital's rule (simplest form)
If u and v are functions with $u(a) = 0$ and $v(a) = 0$, the derivatives $u'(a)$ and $v'(a)$ are defined, and the derivative $v'(a) \neq 0$, then

$$\lim_{x\to a} \frac{u}{v} = \frac{u'(a)}{v'(a)}.$$

We will generalize L'Hôpital's rule in section 6.3, p. 148.

Example 1

▷ Evaluate

$$\lim_{x\to 0} \frac{\sin x}{x + x^3}$$

▷ Attempting to plug in $x = 0$ gives the indeterminate form $0/0$, and this suggests applying L'Hôpital's rule. The derivative of the top is $\cos x$, and the derivative of the bottom is $1+3x^2$. Evaluating these at $x = 0$ gives 1 and 1, so the answer is $1/1 = 1$.

Example 2

The limit

$$\lim_{x\to 1} \frac{3x^2 - x - 2}{x^2 - 1}$$

is of the form $\frac{0}{0}$, so we can try to apply l'Hôpital's rule. We get

$$\lim_{x\to 1} \frac{3x^2 - x - 2}{x^2 - 1} = \frac{6x - 1}{2x} = \frac{5}{2}$$

6.3 Fancier versions of L'Hôpital's rule

Mathematical theorems are sometimes like cars. I own a Honda Fit that is about as bare-bones as you can get these days, but persuading a dealer to sell me that car was like pulling teeth. The salesman was absolutely certain that any sane customer would want to pay an extra $1,800 for such crucial amenities as upgraded floor mats and a chrome tailpipe. L'Hôpital's rule in its most general form is a much fancier piece of machinery than the stripped-down model described in section 6.2. The price you pay for the deluxe model is that the proof becomes much more complicated. I'll state the fancier versions of L'Hôpital's rule below and give examples, but relegate the proofs to a later section and, in one case, a homework problem.

6.3.1 Multiple applications of the rule

In the following example, we have to use l'Hôpital's rule twice before we get an answer.

Example 3

▷ Evaluate

$$\lim_{x\to\pi} \frac{1+\cos x}{(x-\pi)^2}$$

▷ Applying l'Hôpital's rule gives

$$\frac{-\sin x}{2(x-\pi)},$$

which still produces $0/0$ when we plug in $x=\pi$. Going again, we get

$$\frac{-\cos x}{2} = \frac{1}{2}.$$

This works because of the following generalization of L'Hôpital's rule

Theorem: L'Hôpital's rule (first generalization)
If u and v are functions with $u(a)=0$ and $v(a)=0$, and the derivatives $u'(a)$ and $v'(a)$ are defined, then

$$\lim_{x\to a} \frac{u}{v} = \lim_{x\to a} \frac{u'(x)}{v'(x)}.$$

The difference from the original form of the theorem is that we no longer require $v'(a) \neq 0$, and the right-hand side has a limit. In cases where $v'(a) \neq 0$, the original form would have been good enough, but the general form also works, since the limit on the right-hand side can be evaluated simply by plugging in. We will prove this more general form of the rule in section 6.3.4, p. 151.

6.3.2 The indeterminate form ∞/∞

Consider an example like this:

$$\lim_{x\to 0}\frac{1+1/x}{1+2/x}.$$

This is an indeterminate form like ∞/∞ rather than the $0/0$ form for which we've already proved l'Hôpital's rule. L'Hôpital's rule applies to examples like this as well. This can be proved by rewriting an expression like $\lim u/v$, where both u and v blow up, in terms of new variables $U = 1/u$ and $V = 1/v$. The result is to reduce the ∞/∞ form to the $0/0$ form. The proof is carried through in section 6.3.4, p. 151.

Example 4

▷ Evaluate

$$\lim_{x\to 0}\frac{1+1/x}{1+2/x}.$$

▷ Both the numerator and the denominator go to infinity. Differentiation of the top and bottom gives $(-x^{-2})/(-2x^{-2}) = 1/2$. We can see that the reason the rule worked was that (1) the constant terms were irrelevant because they become negligible as the $1/x$ terms blow up; and (2) differentiating the blowing-up $1/x$ terms makes them into the same x^{-2} on top and bottom, which cancel.

Note that we could also have gotten this result without l'Hôpital's rule, simply by multiplying both the top and the bottom of the original expression by x in order to rewrite it as $(x+1)/(x+2)$.

6.3.3 Limits at infinity

It is straightforward to prove a variant of l'Hôpital's rule that allows us to do limits at infinity. We use a change of variable to change a limit like $\lim_{x\to\infty} u(x)/v(x)$ to a new limit stated in terms of a variable $X = 1/x$. The proof is left as an exercise (problem z1, p. 154). The result is that l'Hôpital's rule is equally valid when the limit is at $\pm\infty$ rather than at some real number a.

Acme or Glutco? Example 5

▷ You have some money, and two choices of what to invest it in. A share in Acme, Inc., costs \$7, and returns a dividend of \$1 per year. A share of Glutco costs \$30 and gives a dividend of \$2 per year. If we want to compare the long-term value of the two investments, a natural way to do it is with the limit

$$\lim_{t\to\infty}\frac{-7+t}{-30+2t}.$$

The top represents the net return on Acme, the bottom Glutco. If this limit is greater than 1, then Acme is the better long-term investment. What is the value of this limit?

▹ Differentiation of the top gives 1, and differentiation of the bottom gives 2. The limit is therefore $1/2$, and you're wiser to invest in Glutco. The interpretation is that the constant terms are irrelevant, and in the long run the competition between the numerator and denominator is determined by which one *grows* faster.

6.3.4 Proofs

The simplest form of l'Hôpital's rule was proved in section 6.2, p. 146. In this section we prove the generalizations of l'Hôpital's rule claimed in sections 6.3.1-6.3.3.

Change of variable

As described briefly in sections 6.3.2 and 6.3.3, two of the added features of the generalized l'Hôpital's rule (the form ∞/∞ and limits at infinity) can be proved by a change of variable. To demonstrate how this works, let's imagine that we were starting from an even more stripped-down version of l'Hôpital's rule than the one in section 6.2, p. 146. Say we only knew how to do limits of the form $x \to 0$ rather than $x \to a$ for an arbitrary real number a. We could then evaluate $\lim_{x\to a} u/v$ simply by defining $t = x - a$ and reexpressing u and v in terms of t.

▷ Example 6

Reduce

$$\lim_{x\to\pi} \frac{\sin x}{x - \pi}$$

to a form involving a limit at 0.

▷ Define $t = x - \pi$. Solving for x gives $x = t + \pi$. We substitute into the above expression to find

$$\lim_{x\to\pi} \frac{\sin x}{x - \pi} = \lim_{t\to 0} \frac{\sin(t + \pi)}{t}.$$

If all we knew was the $\to 0$ form of l'Hôpital's rule, then this would suffice to reduce the problem to one we knew how to solve. In fact, this kind of change of variable works in all cases, not just for a limit at π, so rather then going through a laborious change of variable every time, we could simply establish the more general form in section 6.2, p. 146, with $\to a$.

The form ∞/∞

To prove that l'Hôpital's rule works in general for ∞/∞ forms, we do a change of variable on the *outputs* of the functions u and v rather than their inputs. Suppose that our original problem is of the form

$$\lim \frac{u}{v},$$

where both functions blow up.[2] We then define $U = 1/u$ and $V = 1/v$. We now have

$$\lim \frac{u}{v} = \lim \frac{1/U}{1/V} = \lim \frac{V}{U},$$

and since U and V both approach zero, we have reduced the problem to one that can be solved using the version of l'Hôpital's rule already

[2] Think about what happens when only u blows up, or only v.

proved for the indeterminate form 0/0:

$$\lim \frac{u}{v} = \lim \frac{V'}{U'}$$

Differentiating and applying the chain rule, we have

$$\lim \frac{u}{v} = \lim \frac{-v^{-2}v'}{-u^{-2}u'}.$$

Since $\lim ab = \lim a \lim b$ provided that $\lim a$ and $\lim b$ are both defined (property P_5, p. 95), we can rearrange factors to produce the desired result.

Limits at infinity

As briefly outlined in section 6.3.3, this proof can be done by using a change of variables of the form $X = 1/x$. The proof is left as an exercise (problem z1, p. 154).

Problems

a1 Verify the following limits.

$$\lim_{s\to 1}\frac{s^3-1}{s-1}=3$$
$$\lim_{\theta\to 0}\frac{1-\cos\theta}{\theta^2}=\frac{1}{2}$$
$$\lim_{x\to\infty}\frac{5x^2-2x}{x}=\infty$$
$$\lim_{n\to\infty}\frac{n(n+1)}{(n+2)(n+3)}=1$$
$$\lim_{x\to\infty}\frac{ax^2+bx+c}{dx^2+ex+f}=\frac{a}{d}$$

[Granville, 1911] ▷ Solution, p. 238

a2 Evaluate

$$\lim_{x\to 0}\frac{x\cos x}{1-2^x}$$

exactly, and check your result by numerical approximation.
▷ Solution, p. 238

a3 Amy is asked to evaluate

$$\lim_{x\to 0}\frac{x}{e^x}.$$

She applies l'Hôpital's rule, differentiating top and bottom to find $1/e^x$, which equals 1 when she plugs in $x=0$. What is wrong with her reasoning? ▷ Solution, p. 239

a4 Evaluate

$$\lim_{u\to 0}\frac{u^2}{e^u+e^{-u}-2}$$

exactly, and check your result by numerical approximation.
▷ Solution, p. 239

a5 Evaluate

$$\lim_{t\to\pi}\frac{\sin t}{t-\pi}$$

exactly, and check your result by numerical approximation.
▷ Solution, p. 239

d1 Compute the following limits using l'Hôpital's rule.

(a) $\lim_{x\to -1} \frac{x^2-1}{x^2-8x-9}$. ✓

(b) $\lim_{x\to \pi/2} \frac{\sin 2x}{\cos x}$. ✓

(c) $\lim_{x\to 1/2} \frac{\cos \pi x}{1-2x}$. ✓

d2 Suppose n is some positive integer, and the limit

$$\lim_{x\to 0} \frac{\cos x - 1 + x^2/2}{x^n} = L$$

exists. Also suppose $L \neq 0$. What is n? What is the limit L? ✓

d3 What happens when you use l'Hôpital's rule to compute these limits? Compare against what you would have gotten by a more straightforward method.

(a) $\lim_{x\to 0} \frac{x^2}{x}$.

(b) $\lim_{x\to 0} \frac{x^2}{x^3}$.

d4 The logical role of counterexamples was discussed in box 1.3, p. 20. The following rule sounds very much like l'Hôpital's:

if $\lim_{x\to a} \frac{f(x)}{g(x)}$ *exists, then* $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ *also exists, and the two limits are equal.*

But this is not always true! Find a counterexample.

d5 Here is a method for computing derivatives: since, by definition,

$$f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a}$$

is a limit of the form $\frac{0}{0}$, we can always try to find it by using l'Hôpital's rule. What happens when you do that?

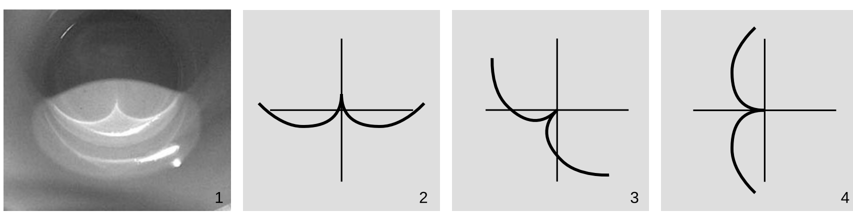
z1 Section 6.3.4, p. 151, demonstrates the use of changes of variable in proving variants on l'Hôpital's rule. As suggested on p. 152, do this for limits at infinity, using the change of variable $X = 1/x$.

Chapter 7

From functions to variables

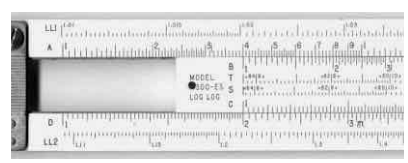
7.1 Some unrealistic features of our view of computation so far

Calculus was invented by Newton and Leibniz, who lived in an era when the best tool for calculation was a freshly sharpened quill, used for writing down formulas. They had in mind a certain *model of computation*. I've introduced you to a related but somewhat different, modern model, based on *functions*. This model doesn't always relate well to reality.



a / Light inside a teacup makes a cusp. Rotating the graph should be irrelevant.

We defined a function geometrically, as a graph that passes the vertical line test. This doesn't work well in an example like figure a. It shouldn't matter whether we take the photo from one angle or another, but if we insist on describing this shape as a function, then rotating it makes a huge difference — the difference between being able to describe the shape and not being able to. In a/2, y is a function of x. In a/3, y isn't a function of x; it fails the vertical line test. In a/4, x is a function of y, but y isn't a function of x. These distinctions are silly in this context. The x and y coordinates are arbitrary, and we shouldn't treat them asymmetrically. We can think of the teacup as a little computer that knows how to compute this particular graph. The teacup doesn't know or care what's x or what's y; neither x nor y is its "input" or "output."



b / This archaic computing device is called a slide rule. Like the teacup in figure a, it's an analog computer, and it doesn't have inputs or outputs. Let A be a number on the scale marked "A," and B the number below it on the "B" scale. Then with the central sliding stick in the position shown in the photo, $A = 4B$.

7.2 Newton's method

In the teacup-computer's personal utopia, there is no distinction between input and output. But if we want to join the teacup in computational nirvana, we have a problem, because we, unlike the teacup, find some functions easier to compute than their inverses. For example, every sixth-grade kid in California is supposed to know how to take the cube of a decimal number such as 4.43. That is, given x, they can compute $y = x^3$. But how many people do you know who can invert the function and efficiently obtain $x = \sqrt[3]{y}$ with paper and pencil? Some functions are computationally cheap to evaluate, but computationally expensive to evaluate in reverse.[1]

Newton, however, invented a method that allows us to at least partially overcome this uninvertibility problem. Newton's method lets us find a good approximation to x for a given y, provided that we know how to evaluate both y and $\mathrm{d}y/\,\mathrm{d}x$ for a given x.

Suppose that we want to find the cube root of 87. We start with a rough mental guess: since $4^3 = 64$ is a little too small, and $5^3 = 125$ is much too big, we guess $x \approx 4.3$. Testing our guess, we have $4.3^3 = 79.5$. We want y to get bigger by 7.5, and we can use calculus to find approximately how much bigger x needs to get in order to accomplish that:

$$
\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &\approx \frac{\Delta y}{\Delta x} \\
\Delta x &\approx \frac{\Delta y}{\mathrm{d}y/\,\mathrm{d}x} \\
&= \frac{\Delta y}{3x^2} \\
&= \frac{\Delta y}{3x^2} \\
&= 0.14
\end{aligned}
$$

Increasing our value of x to $4.3 + 0.14 = 4.44$, we find that $4.44^3 = 87.5$ is a pretty good approximation to 87. If we need higher precision, we can go through the process again with $\Delta y = -0.5$, giving

$$
\begin{aligned}
\Delta x &\approx \frac{\Delta y}{3x^2} \\
&= 0.14 \\
x &= 4.43 \\
x^3 &= 86.9.
\end{aligned}
$$

This second iteration gives an excellent approximation.

[1] An extreme example is embedded in the cryptography systems that allow you to buy something online without worrying that your credit card number is being exposed to random people as it hops across the internet from you to amazon.com. These algorithms depend on the fact that it is computationally cheap to multiply large numbers, but prohibitively expensive to factor a large number into its prime factors.

The orbit of Mercury — Example 1

▷ Figure 1 shows the astronomer Johannes Kepler's analysis of the motion of the planets. The ellipse is the orbit of the planet around the sun. At $t = 0$, the planet is at its closest approach to the sun, A. At some later time, the planet is at point B. The angle x (measured in radians) is defined with reference to the imaginary circle encompassing the orbit. Kepler found the equation

$$2\pi\frac{t}{T} = x - e\sin x,$$

where the period, T, is the time required for the planet to complete a full orbit, and the eccentricity of the ellipse, e, is a number that measures how much it differs from a circle. The relationship is complicated because the planet speeds up as it falls inward toward the sun, and slows down again as it swings back away from it.

The planet Mercury has $e = 0.206$. Find the angle x when Mercury has completed $1/4$ of a period.

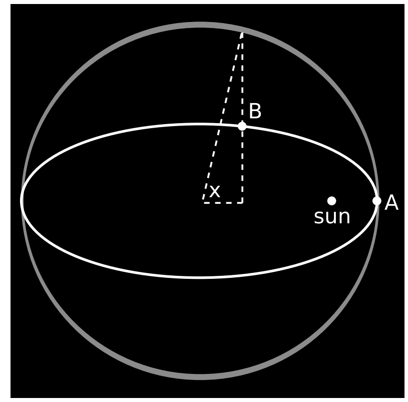
▷ We have

$$y = x - (0.206)\sin x,$$

and we want to find x when $y = 2\pi/4 = 1.57$. As a first guess, we try $x = \pi/2$ (90 degrees), since the eccentricity of Mercury's orbit is actually much smaller than the example shown in the figure, and therefore the planet's speed doesn't vary all that much as it goes around the sun. For this value of x we have $y = 1.36$, which is too small by 0.21.

$$\begin{aligned}\Delta x &\approx \frac{\Delta y}{\mathrm{d}y/\,\mathrm{d}x}\\ &= \frac{0.21}{1-(0.206)\cos x}\\ &= 0.21\end{aligned}$$

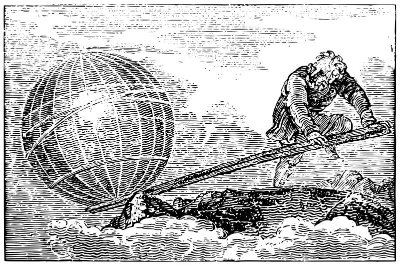
(The derivative $\mathrm{d}y/\,\mathrm{d}x$ happens to be 1 at $x = \pi/2$.) This gives a new value of x, 1.57+.21=1.78. Testing it, we have $y = 1.58$, which is correct to within rounding errors after only one iteration. (We were only supplied with a value of e accurate to three significant figures, so we can't get a result with precision better than about that level.)



c / Example 1.

Usually the series of estimates x_0, x_1, x_2, ... provided by Newton's method *converges*, meaning that $\lim_{n\to\infty} x_n$ exists. Furthermore, the convergence is often very rapid, so that only a few iterations are needed to get excellent precision. But as explored further in problem z1, 171, Newton's method sometimes fails to converge.

7.3 Related rates



d / "Give me a lever and a place to stand, and I will move the world." – Archimedes

Figure d is old and fanciful, but it exemplifies an idea that we use every day. We have some machine or mechanical linkage, which could be as simple as the corkscrew used to open a bottle of wine, or as complicated as the suspension on a fancy sports car. The motion of one part of the machine is not independent of the other parts. In the simple example of a lever, suppose that the heights[2] of the two ends relative to the fulcrum are A on the left and B on the right. Then we have a *constraint* of the form

$$\frac{B}{A} = -k, \tag{1}$$

where k is the ratio of the lengths of the arms, and the minus sign is because if one end goes up, the other has to come down. In figure d, $k \approx 11$; of course Archimedes was imagining k as some very large number, but the cartoonist had to fit everything. Notice that we have no natural reason to call B a function of A or A a function of B. If the arm of the lever is perfectly rigid, then all we can say is that whatever forces act on the ends, the outcome will satisfy the constraint. We don't have to consider one variable as causing the other. (The earth looks more likely to move Archimedes than Archimedes is to move the earth.) In (1), I picked one variable to be on top and the other on the bottom, but instead of $B/A = -11$, I could just as easily have written $A/B = -1/11$.

In examples like this one, we naturally want to know the speed of the motion. How fast will the cork come out of the wine bottle? How fast will my bike go up a hill if I'm in a certain gear? Based on your training so far, you are likely to come up with the following answer for the lever. The position A of the load on the left side of the lever is a function of the position B of the right end, while B is in turn a function of time t. The chain rule therefore gives

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{\mathrm{d}A}{\mathrm{d}B} \cdot \frac{\mathrm{d}B}{\mathrm{d}t}. \tag{2}$$

We know $\mathrm{d}A/\,\mathrm{d}B$, which, based on the constraint, is simply $-1/k$. Next we write down a formula for the function $B(t)$, differentiate it, and plug the result in to equation (2). Done. A triumph of calculus.

Oops. There is no mathematical formula for $B(t)$. The motion of the right end of the lever in figure d comes from an old Greek guy grunting and muttering curses into his white beard.

The term "related rates" is used in calculus to refer to the fact that we don't necessarily *care* whether the function $B(t)$ is known. Often it may be of interest simply to know that *if* B changes at a given rate, *then* A will change at some other rate. These two rates are related to each other by the constraint equation (1).

[2]These heights should actually be measured along circular arcs.

Scuba diving Example 2

When scuba divers ascend or descend, they have to control how fast they go, or else the changes in pressure will be too rapid, and they can be killed. Let P be the pressure in units of atmospheres, y the depth in meters, and t the time in minutes. We then have

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{\mathrm{d}P}{\mathrm{d}y} \cdot \frac{\mathrm{d}y}{\mathrm{d}t}$$

Given the density of water and the strength of the earth's gravity, $\mathrm{d}P/\mathrm{d}y = 0.1$ atm/m. The standard advice is not to ascend faster than $\mathrm{d}y/\mathrm{d}t \approx -10$ m/min. This implies that a diver's body can safely withstand decompression at a rate $\mathrm{d}P/\mathrm{d}t \approx -1$ atm/min.

Cams Example 3

Cams, like the ones shown in figure e, can be thought of as the mechanical realization of the mathematical notion of a function. As the cam rotates, the follower rides up and down above it.

The crankshaft of an engine has its angle φ determined by mechanical linkages (the piston rods) to the pistons. In a four-stroke engine such as the ones in cars, the crankshaft is geared to the camshaft so that the camshaft's angle θ is constrained by $\theta = \varphi/2$. The camshaft then drives each follower, whose height h is controlled by a function $h(\theta)$. This function is determined by the shape of the cam. The followers open and close the valves, which perform functions such as letting fuel into the cylinders. The velocity of the follower is given by

$$\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}\theta} \cdot \frac{\mathrm{d}\theta}{\mathrm{d}\varphi} \cdot \frac{\mathrm{d}\varphi}{\mathrm{d}t},$$

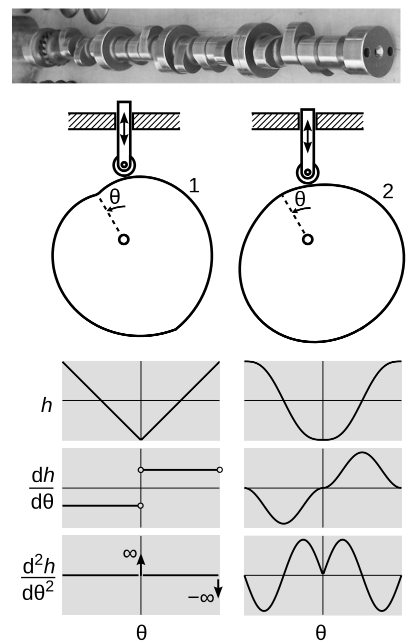
where $\mathrm{d}\varphi/\mathrm{d}t$ is what we measure on a tachometer.

Cam 1 in the figure is shaped so that the follower falls at constant velocity and rises at constant velocity. This has the disadvantage that $\mathrm{d}^2h/\mathrm{d}\theta^2$ is infinite, which would theoretically cause infinite acceleration $\mathrm{d}^2h/\mathrm{d}t^2$ in the follower at the turn-around points. In reality the result would be that the follower would leave contact with the cam, and there would be undesirable vibration.

Cam 2 is shaped according to

$$h(\theta) = 1 + \frac{1}{\pi}\left(|\theta| - \frac{1}{2}\sin(2|\theta|)\right)$$

for $\theta \in [-\pi, \pi]$. This is known as a cycloid cam. It has the desirable property that all of its derivatives up to the third, $\mathrm{d}^3h/\mathrm{d}\theta^3$, are finite, and furthermore that the cycloidal segments of the graph can be joined smoothly onto constant ("dwell") segments without losing these properties. For the reasons discussed in example 5, p. 89, it is desirable not to have a large third derivative.

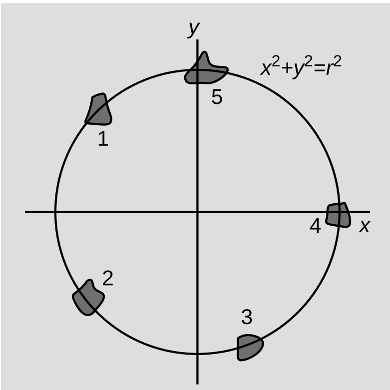


e / Example 3. Top: a racing camshaft from a car. Middle: two cams with specific mathematical shapes. Bottom: Graphs of $h(\theta)$ and its first and second derivatives.

7.4 Implicit functions

As you read this, the world is turning, and you are moving in a circle. Let this circle be centered on the origin, with radius r. Physical forces constrain you to stay on this circle, rather than flying up into the sky or sinking down into the earth's core. The Pythagorean theorem allows us to write this constraint as the equation

$$x^2 + y^2 = r^2, \tag{3}$$



f / The equation $x^2 + y^2 = r^2$ does not define a function unless we restrict it to an appropriate region.

whose solutions are graphed in figure f. This graph fails the vertical line test, so y isn't a function of x, and it also fails the horizontal line test, so x isn't a function of y. Usually by restricting it to a small enough region, we can make it into a function. If we restrict to region 1, 2, 3, or 5, y is a function of x, and similarly for x as a function of y in regions 1, 2, 3, and 4. The largest piece of the graph on which equation (3) defines a function is a semicircle. For example, we could solve for x and find the function

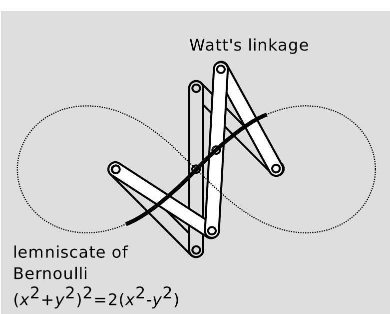
$$x(y) = -\sqrt{r^2 - y^2}, \tag{4}$$

where the choice of the negative square root gives the left-hand half of the circle. Equation (3) is said to define an *implicit* function, while (4) defines an *explicit* one. In an example such as this one, it would be inconvenient to try to work with explicit functions. For example, if we insisted on having explicit functions, we would run into hassles because any calculation would have to be broken down into special cases covering different regions.

Watt's linkage — Example 4

Figure g shows a mechanical linkage patented by James Watt in 1784, and still used in applications such as automobile suspensions. It consists of a chain of three linked rods that are free to rotate about bearings at their ends. The ends of the chain are fixed. The purpose of the arrangement is to constrain some object, attached to the center of the middle rod, to move along the figure-eight curve shown as a dotted line. In this example, the proportions of the three arms are $1 : \sqrt{2} : 1$, so that when the central point is at the center of the curve, they outline a square. This choice of proportions, along with an appropriate choice of scale for the coordinates, can be shown to produce a curve with the equation

$$(x^2 + y^2)^2 = 2(x^2 - y^2). \tag{5}$$



g / Example 4.

In a typical application of a Watt linkage, the central point is attached to the chassis of a car, and the ends are attached to the wheels. The linkage is reoriented so that the darkened segment of the curve is approximately vertical, and the car's chassis is then constrained so that its motion is nearly vertical. When the car goes around corners, the body can't move sideways.

Equation (5) constrains *x* and *y* relative to one another, and makes either variable an implicit function of the other. The linkage can be thought of as a type of computer (an analog computer rather than a digital one) that computes the implicit function (5).

7.5 Implicit differentiation

We would like to be able to do calculus on implicit functions. As a typical application, consider example 4. If vertical motion is desired for small displacements from the center, then we want to rotate the linkage by the correct angle so that the dark portion of the figure-eight curve is vertical near its center. That is, we want to know the slope of the tangent line at this point, so that we can rotate the tangent line and make it vertical. The slope of the tangent line is the derivative, so essentially we need to differentiate a graph that represents an implicit rather than explicit function.

7.5.1 Some simple examples

An example involving addition

But let's start with a simpler example. In figure h, we want to find a proportion between the motion of the tractor and stump. With some arithmetic, we find

$$A + 2B - 2\ell_2 - \ell_1 = 0, \tag{6}$$

which is an implicit relation between A and B. Any change ΔA in the position of the tractor will correspond to some change ΔB in the position of the stump. Setting the change in the left-hand side of equation (6) equal to 0, we have

$$\Delta(A + 2B - 2\ell_2 - \ell_1) = 0.$$

The change in a sum is the same as the sum of the changes, so $\Delta A + 2\Delta B - 2\Delta\ell_2 - \Delta\ell_1 = 0$. But the constants don't change, so

$$\Delta A + 2\Delta B = 0. \tag{7a}$$

The tractor moves twice as much as the stump, and the motion is such that as A increases, B decreases. All of the following are just different ways of expressing the same thought.

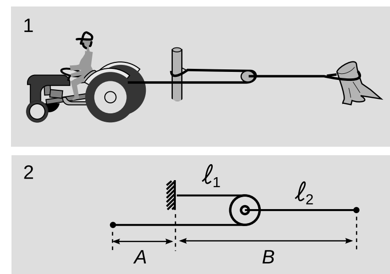
$$\frac{\mathrm{d}A}{\mathrm{d}B} + 2 = 0 \tag{7b}$$

$$1 + 2\frac{\mathrm{d}B}{\mathrm{d}A} = 0 \tag{7c}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} + 2\frac{\mathrm{d}B}{\mathrm{d}t} = 0 \tag{7d}$$

$$\mathrm{d}A + 2\,\mathrm{d}B = 0 \tag{7e}$$

Equation (7e) says that if (7a) works for ordinary numbers like 2 meters and -1 meter, then it should also work for infinitely small numbers (section 2.9, p. 64). Alternatively, some people like to think of an equation like (7e) as nothing more than an informal shorthand for equations involving derivatives such as 7b-7d.



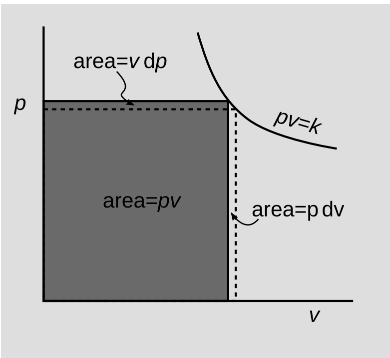
h / 1. Farmer Bill pulls a stump. The pulley is a simple machine, like the lever of section 7.3. Just like the lever, it increases the applied force by some factor, while decreasing the motion by the same factor. 2. In our mathematical model, the fixed post is assumed to be immovable and perfectly rigid, and the ropes perfectly unstretchable, so that their lengths ℓ_1 and ℓ_2 are constant. For simplicity, we neglect the radius of the pulley.

An example with multiplication

Boyle's law states that at a fixed temperature, a sample of an ideal gas has its pressure and volume related by

$$pv = k, \tag{8}$$

where k is a constant. For example, compressing the gas to a smaller volume makes its pressure increase.



i / A geometrical interpretation of equation (9a). Boyle's law says that the areas of the initial, dark rectangle and the final, dashed rectangle are the same. The area $v\,\mathrm{d}p$ lost in the top strip equals the area $p\,\mathrm{d}v$ gained in the side strip.

Suppose that the pressure changes from p to $p + \Delta p$, and the volume from v to $v + \Delta v$. Then:

$$\begin{aligned}
&\Delta(pv) = 0 && \text{[change in each side of (8); } \Delta k = 0\text{]} \\
&(p + \Delta p)(v + \Delta v) - pv = 0 && \text{[subtract initial } pv \text{ from final]} \\
&p\Delta v + v\Delta p + \Delta p\Delta v = 0 && \text{[distribute and cancel } pv \text{ terms]}
\end{aligned}$$

This messy expression can be cleaned up in the case where Δp and Δv are small. The product of two small numbers is even smaller, and if we make them small enough, their product will always be negligibly small compared to them. (Cf. p. 47.) To show that we're now talking about very small numbers, we notate the changes as $\mathrm{d}p$ and $\mathrm{d}v$. We then have:

$$p\,\mathrm{d}v + v\,\mathrm{d}p = 0. \tag{9a}$$

This looks just like the product rule. In this context, symbols like $\mathrm{d}p$ and $\mathrm{d}v$ are referred to as differentials, and we talk about "taking differentials" on both sides of (8) to get (9a). The process of taking differentials is no different than the process of taking a derivative. As in the example of the pulley on p. 161, there are multiple equivalent ways of expressing this statement:

$$p\frac{\mathrm{d}v}{\mathrm{d}p} + v = 0 \tag{9b}$$

$$p + v\frac{\mathrm{d}p}{\mathrm{d}v} = 0 \tag{9c}$$

$$p\frac{\mathrm{d}v}{\mathrm{d}t} + v\frac{\mathrm{d}p}{\mathrm{d}t} = 0 \tag{9d}$$

Some people think of 9a as just a shorthand for (9b)-(9d).

7.5.2 Implicit differentiation in general

Reduced to differentiation of functions

The examples in section 7.5.1 show that no new techniques are needed for implicit differentiation. Every fact about differentiating a function corresponds to a similar fact about implicit differentiation. If we wish, we can do implicit differentiation according to the following recipe, which reduces it to differentiation of a function:

1. Take the equation that defines the implicit function and differentiate both sides with respect to something. It doesn't

matter what we differentiate with respect to; it can be one of the two variables in the equation, or it can be some other variable such as time.

2. (Optional.) If desired, clear all the factors of $1/\,\mathrm{d}\mathit{something}$.

A circle Example 5

▷ The equation $x^2 + y^2 = r^2$ defines a circle. Implicitly differentiate it.

▷ It doesn't matter what we differentiate with respect to, so let's differentiate with respect to t, which lets us imagine that the point (x, y) is moving around the circle as time passes. Since r is a constant, the derivative of the right-hand side is zero.

$$\frac{\mathrm{d}(x^2)}{\mathrm{d}t} + \frac{\mathrm{d}(y^2)}{\mathrm{d}t} = 0$$

Since the expressions x^2 and y^2 aren't written in terms of t, we need to use the chain rule.

$$\frac{\mathrm{d}(x^2)}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\mathrm{d}(y^2)}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}t} = 0$$
$$2x\frac{\mathrm{d}x}{\mathrm{d}t} + 2y\frac{\mathrm{d}y}{\mathrm{d}t} = 0$$
$$x\frac{\mathrm{d}x}{\mathrm{d}t} + y\frac{\mathrm{d}y}{\mathrm{d}t} = 0$$

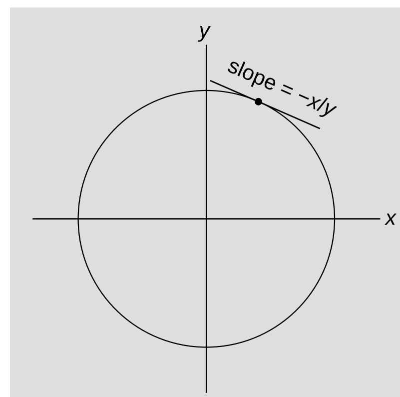
We could stop here if we wished, but the factors of $1/\,\mathrm{d}t$ are messy, and t wasn't even a variable in the original statement of the problem, so it's nicer to multiply by $\mathrm{d}t$ on both sides. We have

$$x\,\mathrm{d}x + y\,\mathrm{d}y = 0 \tag{10}$$

or, equivalently,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{x}{y}. \tag{11}$$

The form (10) has the advantage that it holds anywhere on the circle, whose graph isn't a function. Some people would prefer (11) because they don't believe in Santa Claus or infinitesimals, but it has the disadvantage that it breaks the symmetry between x and y, and it doesn't hold at the two points on the circle where $y = 0$.



j / Examples 5 and 6. The reason for the unexpectedly simple result $\mathrm{d}y/\,\mathrm{d}x = -x/y$ becomes apparent here because the slope of the radius is y/x, and the tangent line must be perpendicular to the radius.

An approximation on the circle Example 6

▷ The following are two nearby points on the unit circle:

$$(0.400000, 0.916515), \qquad (0.401000, 0.916078)$$

Verify that equation (10) is a good approximation.

▷ Since Δx and Δy are small, it makes sense to expect that (10) will be approximately correct if we substitute deltas for the differentials. Let's see if that's true.

$$\begin{aligned} x\Delta x + y\Delta y &= (0.400000)(0.001000) + (0.916515)(-0.000437) \\ &= -0.000001 \end{aligned}$$

The approximation is so good that when we round off to six decimal places, the result almost rounds to zero.

A little bit of ...

Although we saw above that implicit differentiation can be reduced to differentiation of functions, this is not necessary in general. People who are proficient in calculus don't go around making up additional variables like the t in example 5. For example, say that a square has sides of length u. We can think of d as meaning "a little bit of ...,"[3] so that $\mathrm{d}u$ is a little bit of a change in the length of the square's sides. Now u^2 is the area of the square, and $\mathrm{d}(u^2)$ is a little bit of a change in its area. We have a power law that says

$$\mathrm{d}(u^k) = ku^{k-1}\,\mathrm{d}u.$$

This power law is exactly analogous to the one for a function $u(t)$, which, if we apply the chain rule, is

$$\frac{\mathrm{d}(u^k)}{\mathrm{d}t} = ku^{k-1}\frac{\mathrm{d}u}{\mathrm{d}t}.$$

Obviously neither of these needs to be memorized separately from the other. Expressions like $\mathrm{d}u$ and $\mathrm{d}(u^2)$ are known as *differentials.*

Differential of a polynomial — Example 7

▷ Find the differential of $s^2 + s$, and use it to approximate the change in this expression as s changes from 1.000 to 1.001.

▷ For differentiation we have a rule that the derivative of the sum of two functions is the sum of the derivatives. The analogous rule for differentials is that the differential of a sum is the sum of the differentials. Therefore

$$\mathrm{d}(s^2 + s) = \mathrm{d}(s^2) + \mathrm{d}s.$$

Likewise we have a power rule for differentials that corresponds to the power rule for derivatives, and the case of the second power was discussed in detail above. We therefore find

$$\mathrm{d}(s^2 + s) = 2s\,\mathrm{d}s + \mathrm{d}s.$$

The numerical approximation is

$$\Delta(s^2 + s) \approx (2s + 1)\Delta s = (3)(0.001) = 0.003.$$

[3]The phrase is due to the direct and unpretentious Silvanus Thompson, author of a best-selling 1910 calculus textbook.

The power law for fractional exponents Example 8

In section 2.6.3, p. 58, we gave a proof, using only the elementary rules of calculus, that the derivative of $x^{1/2}$ was $\frac{1}{2}x^{-1/2}$, as expected from the power rule. We remarked that although it was clear that such an argument could be constructed for any fractional exponent, that was not the same as giving a general proof. We can write such a proof using implicit differentiation. (We have already proved this fact for any real exponent, using the exponential function, in example 4 on p. 135.)

Let $n = p/q$ where p and q are integers and let

$$y = x^{p/q}.$$

By raising both sides to the power p, we can make this into an implicit function that uses only integer exponents.

$$y^q = x^p.$$

Implicit differentiation gives

$$qy^{q-1}\,\mathrm{d}y = px^{p-1}\,\mathrm{d}x.$$

We then have

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{px^{p-1}}{qy^{q-1}}\\ &= \frac{p}{q}x^{p-1}x^{-(p/q)(q-1)}\\ &= \frac{p}{q}x^{p/q-1}\end{aligned}$$

Example 9

Let $y = f(x)$ be a function defined by

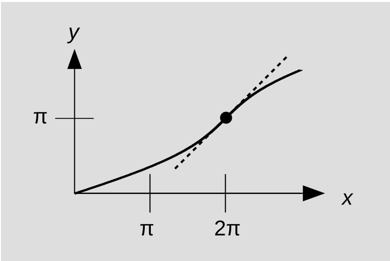
$$2y + \sin y - x = 0.$$

(We encountered a function of this form in a real-world application in example 1, p. 157.) It turns out to be impossible to find a formula that tells you what $f(x)$ is for any given x (i.e., there's no formula for the solution y of the equation $2y + \sin y = x$.) But you can find many points on the graph by picking some y value and computing the corresponding x.

For instance, if $y = \pi$ then $x = 2\pi$, so that $f(2\pi) = \pi$: the point $(2\pi, \pi)$ lies on the graph of f. Let's find how small changes in x and y relate to one another near this point.

Taking differentials on both sides of the defining equation, we have

$$2\,\mathrm{d}y + \cos y\,\mathrm{d}y - \mathrm{d}x = 0$$



k / Example 9. The graph of $x = 2y + \sin y$ contains the point $(2\pi, \pi)$. What is the slope of the tangent line at that point?

or

$$(2 + \cos y)\,dy - dx = 0.$$

We were thinking of y as a function of x. If we wish, we can now find the derivative of this function.

$$\frac{dy}{dx} = \frac{1}{2 + \cos y}$$

If we were asked to find $f'(2\pi)$ then, since we know $f(2\pi) = \pi$, we could answer

$$f'(2\pi) = \frac{1}{2 + \cos\pi} = \frac{1}{2 - 1} = 1.$$

Implicit differentiation was not strictly necessary here, since we could have expressed x as a function of y, found dx/dy, and inverted this to get dy/dx. Our next example is one in which there is no option other than implicit differentiation.

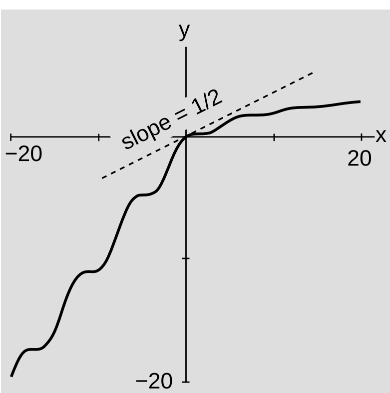
Example 10

▷ Let $x + \cos x = y + e^y$. The graph of this relation passes through the origin. What is its slope there? Check your result numerically with small values of x and y.

▷ We differentiate implicitly.

$$dx - \sin x\,dx = dy + e^y\,dy$$
$$\frac{dy}{dx} = \frac{1 - \sin x}{1 + e^y}$$

Plugging in $x = 0$ and $y = 0$ gives $dy/dx = 1/2$.



l / Example 10. The graph of $x + \cos x = y + e^y$ passes through the origin. Its slope there is $1/2$.

To check this result, we use the approximation $(y - 0)/(x - 0) \approx dy/dx$, which should be valid for small values of x and y. Let's use $x = 0.010$ and $y = 0.005$, which are small and have $y/x = 1/2$, as they approximately should according to the result of our implicit differentiation. If we didn't make a mistake in our calculus, then these values of x and y should be nearly, but not exactly, solutions of the original equation that defined the relation between the variables. Plugging in, we have

$$x + \cos x \stackrel{?}{\approx} y + e^y$$
$$1.00995 \approx 1.01001$$

These are indeed nearly equal, but in fact they were guaranteed to be nearly equal simply because (x, y) was close to the origin, and we knew that the origin was a point on the graph. What we need to check is that the discrepancy between the two sides is small compared to x and y themselves; if $y = (1/2)x$ is the best linear approximation to the graph near the origin, then the error should be on the order of the *squares* of the variables, i.e., something like 10^{-4}. Subtracting, we find that the difference between the two sides of the equation is about 6×10^{-5}, which is indeed small enough to confirm the result of the implicit differentiation.

Implicit differentiation applied to Watt's linkage Example 11

As remarked on p. 161, there is a strong practical motivation for finding the slope of the curve

$$(x^2 + y^2)^2 = 2(x^2 - y^2) \tag{12}$$

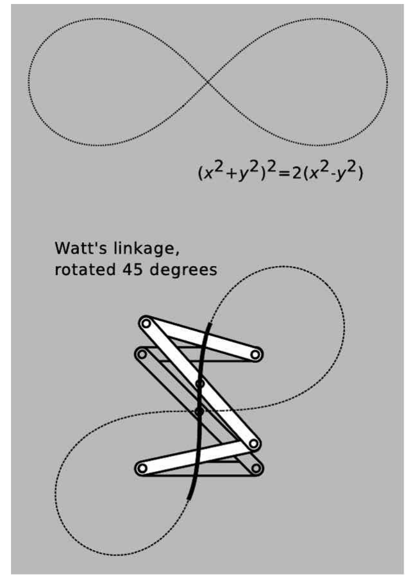
where it passes through the origin. Applying implicit differentiation, we have

$$\begin{aligned}
&2(x^2 + y^2)(2x\ \mathrm{d}x + 2y\ \mathrm{d}y) = 2(2x\ \mathrm{d}x - 2y\ \mathrm{d}y) \\
&(1 - x^2 - y^2)x\ \mathrm{d}x = (1 + x^2 + y^2)y\ \mathrm{d}y \\
&\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{(1 - x^2 - y^2)x}{(1 + x^2 + y^2)y}
\end{aligned}$$

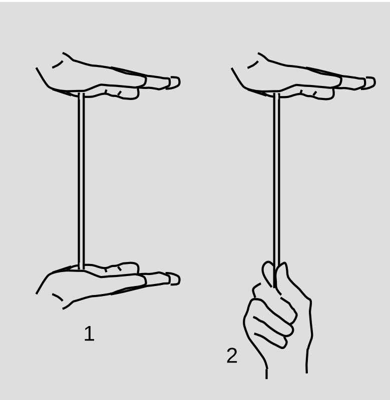
Directly plugging in $x = 0$ and $y = 0$ doesn't work, since this gives $0/0$, which is an indeterminate form (ch. 6). For small values of x and y, the squares x^2 and y^2 become negligible compared to 1, and $\mathrm{d}y/\mathrm{d}x \approx y/x$, so this becomes

$$\begin{aligned}
&\frac{y}{x} \approx \frac{x}{y} \\
&x^2 \approx y^2 \\
&y \approx \pm x.
\end{aligned}$$

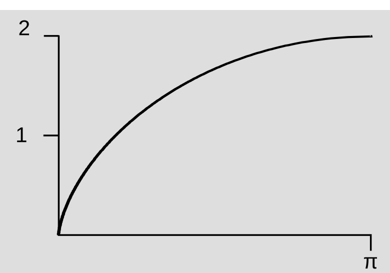
Therefore this curve has a slope of ± 1 on its two segments crossing the origin. To make Watt's linkage (with arms in the proportions previously described) constrain its central point to nearly vertical motion, we need to rotate it by 45 degrees.



m / Example 11.



Problem a1.



Problem a3.

Problems

a1 Figure n/1 shows a thin stick being compressed between a person's hands. If the force is greater than a certain amount, the stick will start to bow. Figure n/2 is similar, but at the bottom the stick is constrained so that it can't rotate; that is, its tangent is kept vertical. The stick is stronger in this situation, and more force is required before it will start to deform. The ratio of the two forces can be shown to be $(x/\pi)^2$, where x is the smallest positive solution of the equation

$$\tan x = x.$$

Inspection of a graph of the tangent function shows that the value of x is approximately 4.5. Use Newton's method to improve this approximation to six decimal places. ✓

a2 The British economist Robert Malthus (1766-1834) theorized that the human population would tend to grow exponentially with time, whereas the production of resources such as food would grow only linearly, due to factors such as technological improvements. Under these assumptions, the population would then inevitably become too great to be fed, resulting in an event now known as a Malthusian catastrophe, such as famine or genocide. As an example, suppose that the production of food in a certain country increases so that at time $t \geq 0$, agriculture can feed a population $2+t$ (in units of millions of people), while the population (in the same units) equals e^t. A Malthusian catastrophe will then occur at a time t determined by

$$2+t=e^t.$$

Use Newton's method to determine t to two decimal places. ✓

a3 The cycloid, figure o, was introduced briefly in example 3, p. 159. It is the shape traced out in space by a point on the rim of a rolling wheel (which in this problem we take to have radius 1). Its equation in Cartesian coordinates can be written as

$$x = \cos^{-1}(1-y) - \sqrt{y(2-y)},$$

which can't be solved for y in terms of x (in the sense defined in section 9.3). Use Newton's method to find the value of y corresponding to $x=1$, expressing your answer to five decimal places. ✓

c1 A sugar cube dissolves in hot tea so that the edge of the cube decreases at a rate $r = \mathrm{d}\ell/\,\mathrm{d}t$. (a) How fast is the volume V of the cube changing when the edge has length ℓ? (b) Check that your answer has units that make sense. (c) Evaluate your answer numerically for $\ell = 5.0$ mm and $r = -0.3$ mm/s (millimeters per second). ✓

c2 (a) A conical water tank with vertex down has height h, and radius a at the top. The water is being drained out at a rate of flow $F = \mathrm{d}V/\,\mathrm{d}t$. How fast is the depth d of the water decreasing, when d has a certain value? (b) Check that your answer to part a has units that make sense. (c) Evaluate your answer numerically for $a = 12$ m, $h = 30$ m, $d = 20$ m, and $\mathrm{d}V/\,\mathrm{d}t = -1.4 \times 10^{-2}$ m^3/s. ✓

Huge talus cones on the coast of Svalbard, problem c3.

c3 The photo shows a common geological formation called *talus*. Erosion causes rock and sand to be washed down the gullies, where over geological time this debris piles up higher and higher against the vertical cliff. Suppose that the pile is in the shape of half a cone, and that its volume grows at a rate $R = \mathrm{d}V/\,\mathrm{d}t$. The cone's slope α is fixed by the maximum steepness for which friction is capable of keeping a rock from sliding down. (a) Find the rate $\mathrm{d}h/\,\mathrm{d}t$ at which the height of the cone grows, in terms of R, α, and h. (b) Check that your answer to part a has units that make sense. (c) Check the dependence of your answer on the variable R. That means that you should determine *physically* whether increasing R should increase the result or decrease it, and then compare this to the *mathematical* behavior of your equation. (d) Do the same for the variables α and h. ✓

c4 During chemotherapy, the volume of a spherical tumor decreases at a rate that is proportional to its surface area. Show that its radius decreases at a constant rate.

In problems e1-e9, evaluate the differentials.

e1 $\mathrm{d}(B^{52})$ ▷ Solution, p. 239

e2 $\mathrm{d}(2000BC)$ ▷ Solution, p. 239

e3 $\mathrm{d}(\sin k)$ ▷ Solution, p. 239

e4 $\mathrm{d}(pb + j)$ ▷ Solution, p. 239

e5 $\mathrm{d}(e^w)$ ✓

e6 $\mathrm{d}(uck)$ ✓

e7 $\mathrm{d}(e^n y)$ ✓

e8 $\mathrm{d}\left(\frac{1}{u^3}\right)$ ✓

e9 $\mathrm{d}(\pi r^2)$ (differential of the area of a circle) ✓

In problems g1-g4, a function $y(x)$ is defined explicitly. Find an implicit definition that does not involve taking roots. Then use this description to find the derivative $\mathrm{d}y/\,\mathrm{d}x$ in terms of x.

g1 $y = \sqrt{x^2 + 1}$ ▷ Solution, p. 239

g2 $y = \sqrt{1 - x}$ ✓

g3 $y = \sqrt[4]{x + x^2}$ ✓

g4 $y = \sqrt{1 - \sqrt{x}}$ ✓

In each of the problems i1-i4, an implicit relation is defined between x and y, and the graph passes through the origin. Find the slope of the graph at the origin.

i1 $xe^{x+y} + y = 0$ ▷ Solution, p. 240

i2 $\sin x - y\cos(xy) = 0$ ✓

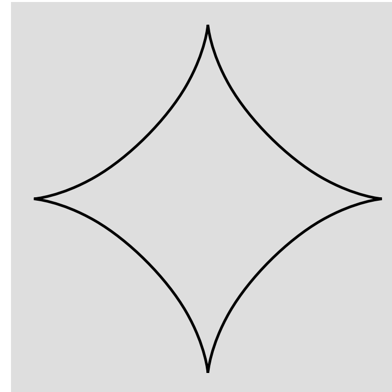
i3 $(x + 2y - 1)^2 + (4x - y - 1)^3 = 0$ ✓

i4 $\sin(ye^x) + e^{x\cos y} - 1 = 0$ ✓

k1 An astroid is the shape traced by a point on a circle of radius $a/4$ as it rolls around the inside of a circle of radius a. Its equation is

$$x^{2/3} + y^{2/3} = a^{2/3}.$$

(a) Check that the units of the equation make sense. (b) Use implicit differentiation to find an extremely simple expression for $\mathrm{d}y/\,\mathrm{d}x$ in terms of y and x. (Do not eliminate y in favor of x, because that makes the expression more complicated.) (c) Check the units of your result. (d) Check that the sign of your result is correct in all four quadrants of the graph. (e) The notion of a *cusp* was briefly introduced on p. 61; it is a horn-shaped point on a graph where the two branches are parallel when they meet at the tip. From the figure, it's hard to tell whether the astroid has cusps or whether there is a nonzero angle between the branches. Use your result to determine which is the case.



An astroid, problem k1.

k2 The figure shows a fountain in Sergel's Square, Stockholm, named after the sculptor Sergel. The fountain was designed by architect David Helldén using a mathematical shape suggested by his friend, the Danish mathematician, poet, designer, and author Piet Hein. The equation of the shape is

$$|x|^{5/2} + |y|^{5/2} = a^{5/2},$$

where a is a constant. (a) Find the units of a. (b) Use implicit differentiation to find an extremely simple expression for $\mathrm{d}y/\,\mathrm{d}x$ in terms of y and x. For simplicity, you can restrict your result to the first quadrant. (Do not eliminate y in favor of x, because that makes the expression more complicated.) (c) Check that the units of your result make sense. (d) Check that the sign of your result makes sense. (e) Check that the result makes sense where the curve intersects the positive x and y axes.

Problem k2.

k3 Evaluate $\mathrm{d}(x^y)$, and show that you can recover the correct results in the special cases where x or y is constant. Hint: rewrite the expression in terms of the exponential function.

z1 Newton's method fails in some cases. As an example, suppose we have $f(x) = |x|^{1/4}$, we want to find an x such that $f(x) = 0$, and we start with $x_0 = 1$ as our initial guess. Of course this is a silly application, since it's obvious that the solution is $x = 0$, but the point is to study a simple example where the method fails. Find a formula for $|x_n - x_{n-1}|$ in this example. Then use this result in a proof by induction to show that Newton's method fails.

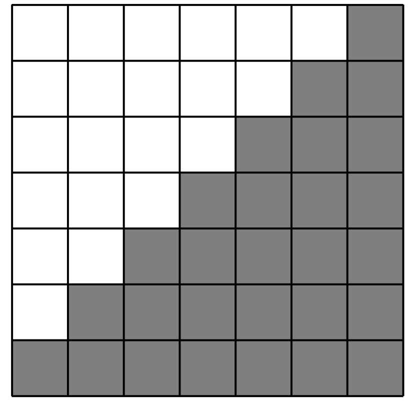
Chapter 8
The integral

8.1 The accumulation of change

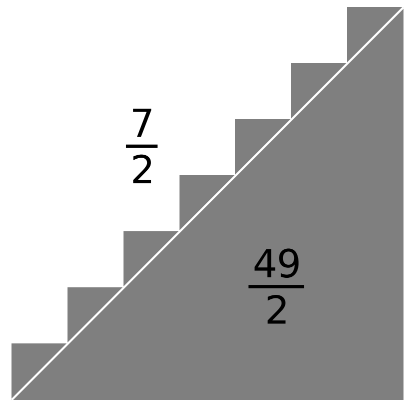
8.1.1 Change that accumulates in discrete steps

A schoolboy plays a trick

Toward the end of the eighteenth century, a German elementary school teacher decided to keep his pupils busy by assigning them a long, boring arithmetic problem: to add up all the numbers from one to a hundred.[1] The children set to work on their slates, and the teacher lit his pipe, confident of a long break. But almost immediately, a boy named Carl Friedrich Gauss brought up his answer: 5,050.



a / Adding the numbers from 1 to 7.



b / A trick for finding the sum.

Figure a suggests one way of solving this type of problem. The filled-in columns of the graph represent the numbers from 1 to 7, and adding them up means finding the area of the shaded region. Roughly half the square is shaded in, so if we want only an approximate solution, we can simply calculate $7^2/2 = 24.5$.

But, as suggested in figure b, it's not much more work to get an exact result. There are seven sawteeth sticking out out above the diagonal, with a total area of $7/2$, so the total shaded area is $(7^2 + 7)/2 = 28$. In general, the sum of the first n numbers will be $(n^2 + n)/2$, which explains Gauss's result: $(100^2 + 100)/2 = 5,050$.

There is a tantalizing hint here of a link with differential calculus, because the derivative of a real function $f(n) = (n^2+n)/2$ is almost, but not quite, equal to n.

c / Carl Friedrich Gauss (1777-1855), a long time after graduating from elementary school.

Accumulation of change in discrete steps

Problems like this come up frequently. Imagine that each household in a certain small town sends a total of one ton of garbage to the dump every year. Over time, the garbage accumulates in the dump, taking up more and more space. If the population is constant, then garbage accumulates at a constant rate. But maybe the town's population is growing. If the population starts out as 1 household in year 1, and then grows to 2 in year 2, and so on, then we have the same kind of problem that the young Gauss solved. After 100 years, the accumulated amount of garbage will be 5,050 tons. The

[1] I'm giving my own retelling of a hoary legend. We don't really know the exact problem, just that it was supposed to have been something of this flavor.

pile of refuse grows more quickly every year.

Sigma notation

There is a convenient way of notating sums like the ones we've been doing, which involves Σ, called "sigma," the capital Greek letter "S." Here the "S" stands for "sum." The sigma notation looks like this:

$$\sum_{i=1}^{100} i = 5,050 \qquad (1)$$

This is read as "the sum of i for i from 1 to 100 equals 5,050." The version without the sigma notation is much more cumbersome to write:

$$1 + 2 + 3 + \ldots + 100 = 5,050 \qquad (2)$$

In equation (1), i is a dummy variable. We could have written

$$\sum_{j=1}^{100} j = 5,050$$

and it would have meant exactly the same thing. We've already seen some examples of dummy variables. In set notation (box 1.1, p. 15),

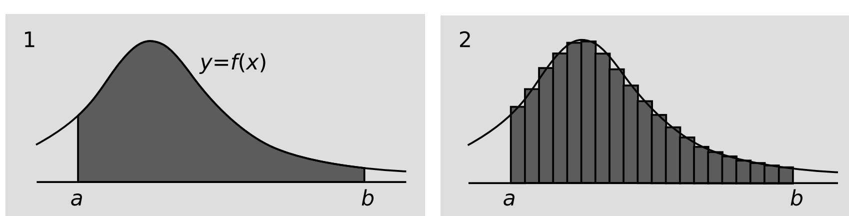
$$\mathrm{S} = \{x|x^2 > 0\} \quad \text{and} \quad \mathrm{T} = \{y|y^2 > 0\}$$

describe exactly the same set, and S=T. Similarly, the function f defined by $f(u) = u^2$ and the function g defined by $g(v) = v^2$ are the same function, $f = g$.

8.1.2 The area under a graph

The examples in section 8.1.1 involved change that occurred in discrete steps. Calculus is concerned with *continuous* change. The continuous analog of a discrete sum is the area under a graph. Let f be a function that is defined on an interval[2] $[a, b]$ and assume the value of f is always positive (so that its graph lies above the x axis). *How large is the area of the region caught between the x axis, the graph of $y = f(x)$ and the vertical lines $y = a$ and $y = b$?*

d / Bernhard Riemann (1826-1866).



e / 1. The area under the graph of the function f. 2. Approximating this area using 20 thin rectangles.

[2]For interval notation, see p. 15.

8.1.3 Approximation using a Riemann sum

We can try to compute this area, figure e1, by approximating the region with many thin rectangles, e2. The idea is that even though we don't know how to compute the area of a region bounded by arbitrary curves, we do know how to find the area of one or more rectangles. In this example, we've subdivided the interval from a to b into $n = 20$ equal subintervals, each of width $\Delta x = (b - a)/n$. Let's write x_1 for the x value that lies in the center of the first subinterval, etc. We've chosen the height of each rectangle so that its top intersects the graph at this midpoint, so that, e.g., the height of the first rectangle is $f(x_1)$. The area of the k^{th} rectangle is the product of its height and width, which is $f(x_k)\Delta x$. Adding up all the rectangles' areas yields

$$R = \sum_{k=1}^{n} (\text{height})(\text{width}) = \sum_{k=1}^{n} f(x_k)\Delta x. \tag{3}$$

This is an example of what is called a *Riemann sum*, meaning an approximation to the area under a curve using rectangles. This particular type of Riemann sum is one in which (a) the interval is subdivided into equal parts, and (b) the value of the function is sampled at the center of each subinterval.

If f is negative in certain places, then we will hit certain values of k for which the product $f(x_k)\Delta x$ is negative. We will simply *define* areas below the x axis to be negative. We think of the rectangle as having positive width Δx but negative height $f(x_k)$. A similar geometrical example is the use of negative numbers for angles that are directed contrary to a standard direction of rotation.

If our rectangles are all sufficiently narrow then we expect the total area of all the rectangles to be a good approximation of the area of the region under the graph.

8.2 The definite integral

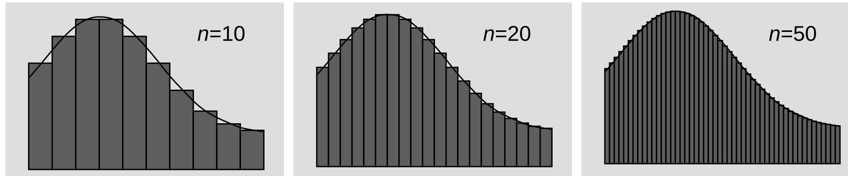
8.2.1 Definition of the integral of a continuous function

This suggests the following definition.

Definition of the integral of a continuous function
If f is a continuous function defined on an interval $[a, b]$, then the integral of $f(x)$ from $x = a$ to b is defined as

$$\lim_{\Delta x \to 0} R,$$

where R is the type of Riemann sum defined above, using equal subintervals sampled at their centers.

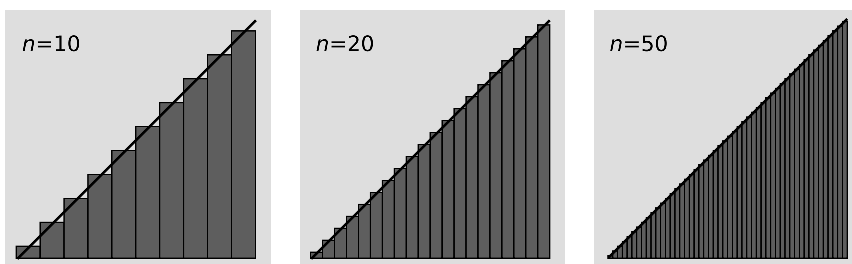


f / Three Riemann sums for the same function on the same interval. As Δx approaches zero, the total area approaches the Riemann integral.

Finding the integral of a function referred to as integrating it. The idea behind the words is that one meaning of "integrate" in ordinary speech is to assemble a whole out of smaller parts. For example, you could integrate sit-ups into your routine at the gym.

Up until now we've been doing differential calculus. The other half of calculus, integral calculus, consists of the study of integrals. The type of integral defined here is called a definite integral. We'll see later that there is another type, called the indefinite integral.

This definition is restricted to continuous functions. A more general definition is given in section 8.6.2, p. 192.



g / Example 1.

A triangle Example 1

Let $f(x) = x$. Then the integral of f from 0 to 1 represents the area of a triangle with height 1 and a base of width 1. We know from elementary geometry that this shape has an area equal to $\frac{1}{2}$(base)(height) = $\frac{1}{2}$, so we don't need integral calculus to determine it. But let's see how this works out if we do it as an integral, in order to get comfortable with the tool and see if it works in a case where we already know the answer.

When we split up the interval $[0,1]$ into n parts, we have $\Delta x = 1/n$. The first subinterval is $[0,\Delta x]$, and its center is the first sample point, $x_1 = (1/2)\Delta x$. Continuing in this way, we have $x_k = (k - 1/2)\Delta x$, for k running from 1 to n. Since our function is just $f(x) =$

x, we also have $f(x_k) = (k - 1/2)\Delta x$. The Riemann sum R is shown in figure g. It looks almost exactly like the staircase in a on p. 173. There are two differences: (1) in the original staircase problem, the graph covered a region of graph paper n squares wide and n squares tall, whereas the graph of our Riemann sum is scaled down so that it fits inside a single square with a width of 1 and a height of 1; (2) all of the steps have been lowered by half a step.

When we evaluate the Riemann sum, we find that the fates have been kind to us, and its value in this example always seems to be 1/2, for every n. For example, with $n = 3$ the Riemann sum is $\frac{1}{6}\Delta x + \frac{1}{2}\Delta x + \frac{5}{6}\Delta x = \frac{9}{6}\Delta x = \frac{1}{2}$.

To see that this is always true in this example, let's go ahead and compute the Riemann sum for an arbitrary n.

$$\begin{aligned}
R &= \sum_{k=1}^{n} f(x_k)\Delta x \\
&= \sum_{k=1}^{n} \left[\left(k - \frac{1}{2}\right)\Delta x\right]\Delta x \\
&= (\Delta x)^2 \sum_{k=1}^{n} \left(k - \frac{1}{2}\right) \\
&= (\Delta x)^2 \left[\left(\sum_{k=1}^{n} k\right) - \left(\sum_{k=1}^{n} \frac{1}{2}\right)\right] \\
&= (\Delta x)^2 \left[\left(\sum_{k=1}^{n} k\right) - \frac{n}{2}\right]
\end{aligned}$$

The sum over k is the same one that we encountered in our previous study of the "staircase" sum; it equals $(n^2 + n)/2$. The result is:

$$\begin{aligned}
R &= (\Delta x)^2 \left\{\left[\frac{n^2 + n}{2}\right] - \frac{n}{2}\right\} \\
&= (\Delta x)^2 \frac{n^2}{2}
\end{aligned}$$

But $\Delta x = 1/n$, so $R = 1/2$ exactly for every n, and the integral equals

$$\lim_{n\to\infty} R = \frac{1}{2},$$

as expected geometrically.

8.2.2 Leibniz notation

If we take equation (3) that defines the Riemann sum R, and substitute it into the definition of the integral $\lim_{\Delta x \to 0} R$, the result looks like this:

$$\lim_{\Delta x \to 0} \sum_{k=1}^{n} f(x_k)\Delta x$$

Leibniz invented the following expressive, versatile, and useful notation for this limit:

$$\int_a^b f(x)\,\mathrm{d}x$$

The symbol $\int$ is an "S" that's been stretched like taffy. It stands for "sum," just as the sigma, Σ, stands for "sum." But we think of $\int$ as meaning a *smooth* sum, whose graphical representation is the area under a smooth curve rather than under a staircase. Notice how the shape of $\int$ is smooth. Like the k in the sigma notation, the x in this example is a dummy variable. Therefore $\int_a^b f(x)\,\mathrm{d}x$ means exactly the same thing as $\int_a^b f(s)\,\mathrm{d}s$. The dummy variable inside an integral is referred to as a variable of integration, and has no meaning outside the integral. One of the reasons for writing the $\mathrm{d}x$ is that it states what we're integrating with respect to.

Leibniz notation for the area of a triangle — Example 2

In example 1, we integrated the function $f(x) = x$ from $x = 0$ to 1, and found that it was 1/2. In Leibniz notation, the result is written like this:

$$\int_0^1 x\ \mathrm{d}x = \frac{1}{2}$$

It makes no difference if we notate this instead with s as the variable of integration:

$$\int_0^1 s\ \mathrm{d}s = \frac{1}{2}$$

A rectangle — Example 3

▷ Evaluate

$$\int_0^4 1\ \mathrm{d}x.$$

▷ The graph of this function is a rectangle with height 1 and width 4. A rectangle is a shape that can be sliced up into thin, vertical slices that are also rectangles, and this is what any Riemann-sum approximation to this integral will look like. The approximations aren't really approximations at all. Every Riemann sum has an area of 4, so the limit occurring in the definition of the integral is 4. This is of course the correct result for the area of this rectangle.

We defined the Leibniz notation as simply a notation for a certain limit, but we can think of it conceptually as a sum with infinitely many terms. That is, we make a Riemann sum with infinitely many rectangles. Normally if you added up an infinite number of things, you would expect to get an infinite result. But remember, each of these rectangles is infinitely skinny. We think of $\mathrm{d}x$ as being the infinitely small width, so that the area $f(x)\,\mathrm{d}x$ is infinitely small. We're therefore adding an infinite number of things, each of which is infinitely small, so that the result can be finite. Recall that, as discussed in section 2.9, p. 64, the real number system doesn't have infinitely big or infinitely small numbers; however, if we handle our infinities according to the simple rules given in that section, nothing bad happens. Historically, these rules weren't formalized, and practitioners just knew that if they did their work according to certain methods, the Leibniz notation never led to the wrong result. This confusion was definitively cleared up around 1965, but many mathematicians have been influenced by the historical uneasiness about the Leibniz notation, so they prefer to think of $\int \ldots \mathrm{d}x$ purely as a shorthand notation for a limit. This is a matter of taste. Those who prefer to think of it only as a shorthand will consider the $\mathrm{d}x$ inside the integral to be nothing more than punctuation, like the period at the end of a sentence. From this point of view, its only job is to tell us what the dummy variable is, i.e., what we're integrating with respect to.

Moving the dx *around* — Example 4

One of the rules in section 2.9 was that we were allowed to manipulate differentials such as $\mathrm{d}x$ using any of the elementary axioms of the real numbers (section 1.6, p. 25). One of these axioms is that multiplication is commutative, $uv = vu$. Therefore the integral in example 2 on p. 178 can be written in either of the following equivalent ways:

$$\int_0^1 x\ \mathrm{d}x = \frac{1}{2} \qquad \int_0^1 \mathrm{d}x\ x = \frac{1}{2}$$

Similarly, all of the following are the same integral:

$$\int_1^2 \frac{1}{x}\ \mathrm{d}x = \int_1^2 \mathrm{d}x\ \frac{1}{x} = \int_1^2 \frac{\mathrm{d}x}{x}$$

Most people would write it with the $\mathrm{d}x$ on top, which makes it more compact.

The integral of . . . what? — Example 5

How should we interpret this expression?

$$\int_0^4 \mathrm{d}x$$

There doesn't seem to be any function written inside the integral, so what is it that we're integrating? One of the elementary axioms of the real numbers (section 1.6, p. 25) is that 1 is the multiplicative identity, i.e., $1u = u$ for any number u. As discussed in section

2.9, the elementary axioms also apply to differentials. Therefore it's valid to rewrite our integral as follows.

$$\int_0^4 1\,\mathrm{d}x$$

The function we're integrating is 1, which makes this the same integral as the one in example 3 on p. 178. The result is 4.

Another way of interpreting the original form of the integral is that d*x* means "a little bit of *x*," so that the integral expresses the idea of letting *x* change from 0 to 4, and adding up all the little changes in *x*. Clearly the sum of all the little changes will be the total change, which is 4.

Another nice feature of the Leibniz notation is that it makes the units come out right. Consider our earlier example of the town dump. Suppose that the rate of garbage production is given by a function $p(t)$, where t is in units of years and p in tons per year. Then the amount of garbage accumulated at the town dump from year a to year b is given by

$$\int_a^b p(t)\ \mathrm{d}t.$$

The integral sign $\int$ is a kind of sum, and the units of a sum are the same as the units of each term. Since d means "a little bit of ...," dt stands for a little bit of time, and it therefore also has units of years. The units of the terms in the sum are

$$\frac{\text{tons}}{\text{year}} \times \text{years} = \text{tons},$$

which makes sense.

We can now see three independent reasons why an integral such as $\int_0^1 x^2\ \mathrm{d}x$ can't be written like $\int_0^1 x^2$, without the dx:

1. If x has units, then the expression without the dx has the wrong units.

2. It would be a sum of infinitely many numbers, each of them finite, so it would probably be infinite.

3. If we don't write the dx, we haven't stated what we're integrating with respect to.

8.3 The fundamental theorem of calculus

8.3.1 A connection between the derivative and the integral

We've already seen some clear indications of a link between derivatives and integrals. A derivative is a rate of change, and an integral measures the accumulation of change. Let's say for concreteness that we're talking about functions of time. If a function A tells us the rate at which function B changes, then B tells us how the rate of change measured by A has accumulated over time. That is, it seems clear conceptually that the integral and the derivative are inverse operations: operations that undo each other, in the same way that subtraction undoes addition, or a square root undoes a square.

	A	B
1	garbage per year	accumulated garbage
2		0
3	1	1
4	2	3
5	3	6
6	4	10
7	5	15
8	6	21
9	7	28
10		

h / Columns A and B in the spreadsheet relate to each other approximately as derivative and integral.

Figure h shows this in the context of discrete rather than continuous functions. Column A shows how many tons of garbage are sent to the town dump per year. It is the rate of change of the pile at the dump, which is given in column B. The population is growing, so column A is not constant. Presumably one of these columns was typed into the spreadsheet from data collected by the town, but we can't tell from looking at the spreadsheet which one it was. It's possible that the raw data was column A, in which case column B would have been constructed by telling the spreadsheet software to calculate a running sum based on A. The running sum of a discrete function is conceptually similar to the integral of a continuous one, so we can say that in some loose sense that B is the integral of A. On the other hand, it's possible that the raw data was column B: a municipal employee has been going out to the dump at yearly intervals and measuring how big the pile of trash was. Column A would then have been calculated from B by taking differences of successive years. This is conceptually similar to saying that A is the derivative of B.

8.3.2 What the fundamental theorem says

The fundamental theorem of calculus

Let f be a function defined on the interval $[a, b]$, and let f be differentiable on that interval. Then

$$\int_a^b \frac{\mathrm{d}f}{\mathrm{d}x}\,\mathrm{d}x = f(b) - f(a). \tag{4}$$

On the left-hand side, we have taken a function, differentiated it, and then integrated it. The right-hand side is a simple expression involving the original function, i.e., in some sense the integration has undone the differentiation, and we are left with the same function we started with.

	A	B
1	garbage per year	accumulated garbage
2		1000
3	1	1001
4	2	1003
5	3	1006
6	4	1010
7	5	1015
8	6	1021
9	7	1028

i / The initial amount of garbage is 1000 tons rather than zero.

To see why the right-hand side contains a difference of two values of f, consider figure i, which is a modified version of h. What's changed is that rather than starting out empty in the first year,

j / After translation by a computer from English to Chinese, and then back to English, the original sentence is not quite the same. By analogy, the fundamental theorem tells us that if we differentiate, then integrate, we cannot quite recover the original function: we lose any information that amounts to an over-all additive constant.

in this version of history the dump started out with 1000 tons of garbage already in it. This alteration of column B, however, has no effect on column A. For example, the subtraction $1015 - 1010$ gives the same result as $15 - 10$. The fundamental theorem tells us that we can make a "round trip" by computing column A from column B using differences, and then reconstructing column B again by taking a running sum. But the round trip isn't perfect (cf. figure j). Some information is lost, because given column A, we can't tell whether the version of column B we should reconstruct is the one in figure h, the one in i, or some other version that differs from them by some other additive constant. What we *can* tell is that the *difference* between the initial and final cells of column B must have been 28, which is the sum of column A.

In terms of continuous functions rather than discrete ones, adding a constant onto f doesn't change the derivative $\mathrm{d}f/\,\mathrm{d}x$. Therefore the left-hand side of the fundamental theorem can never tell us the *value* of f but only the *difference in values* between $x = a$ and $x = b$.

8.3.3 A pseudo-proof

We've seen examples before in which the Leibniz notation makes certain facts about calculus seem so obvious that they don't seem to need any further proof. This happens, for example, if we rewrite the chain rule as $\mathrm{d}z/\,\mathrm{d}x = (\mathrm{d}z/\,\mathrm{d}y)(\mathrm{d}y/\,\mathrm{d}x)$, which makes it seem like a simple fact about algebra; but this is not quite a rigorous proof for the reasons explained in example 18, p. 66. It's a "pseudo-proof," but that's not necessarily a bad thing. Pseudo-proofs can be good. The pseudo-proof helps us to understand why the result makes sense, and it can, if we wish, serve as the backbone of a more rigorous proof.

We will give a real proof of the fundamental theorem in section 8.6.3, p. 194, but let's warm up with the pseudo-proof, which is pretty simple. We start with a statement of the result,

$$\int_a^b \frac{\mathrm{d}f}{\mathrm{d}x}\,\mathrm{d}x \stackrel{?}{=} f(b) - f(a), \tag{5}$$

with the question mark above the equals sign to show that this is what we are hoping to prove. For the same reasons as in example 18 on p. 66, it is not quite valid to cancel the factors of $\mathrm{d}x$, but we'll do it anyway because this is only meant to be a pseudo-proof.

$$\int_{f(a)}^{f(b)} \mathrm{d}f \stackrel{?}{=} f(b) - f(a) \tag{6}$$

We can interpret the symbol $\mathrm{d}f$ as "a little bit of f," so that the left-hand side is the sum of many very small changes in f. The limits of integration are now stated in terms of the values of f, since f is now the variable of integration, not x. (It's true, but not as obvious, that this is equally valid regardless of whether f is always increasing or always decreasing. If f goes up and then comes back down, we could, for example, have $f(a) = f(b)$, so that the upper and lower limits of integration were the same.)

It's clearly reasonable now to hope that we can make the left-hand side of equation (6) equal the right. The left-hand side says that we add up many small changes in the variable f. The right-hand side is simply the total accumulated change in f. To see this a little more explicitly, let's insert a factor of 1 inside the integral, as in example 5, p. 179.

$$\int_{f(a)}^{f(b)} 1 \cdot \mathrm{d}f = f(b) - f(a) \tag{7}$$

As in that example, this integral represents the area of a rectangle. The rectangle has width $f(b) - f(a)$ and height 1, so its area is $f(b) - f(a)$, and the equation holds.

This pseudo-proof is refined into a real proof in section 8.6.3, p. 194,

8.3.4 Using the fundamental theorem to integrate; the indefinite integral

Avoiding the Riemann sum

The fundamental theorem says this:

$$\int_a^b f'(x)\,\mathrm{d}x = f(b) - f(a).$$

In some examples, this gives us a tricky way to evaluate an integral exactly without having to muck around with Riemann sums. Consider the integral

$$\int_0^1 x\ \mathrm{d}x,$$

whose geometrical interpretation is the area of a triangle and whose value we showed to be 1/2 using Riemann sums in example 1, p. 176. The function we're integrating is x, but what if we could find a function f whose derivative was x? —

$$f'(x) = x$$

The fundamental theorem would then immediately tell us the result of the integral.

Antiderivatives

The function f is called an *antiderivative* of the function f'. Although there are various tricks and methods for finding antiderivatives, in general the only way to find them is to guess and check. One way to approach this one is to think of x as x^1. We know that when we differentiate a power, the power rule tells us to knock down the exponent by one. That makes it reasonable to guess something like x^2 as an antiderivative of x. Checking our guess, we find that it was almost, but not quite, right:

$$f(x) = x^2 \implies f'(x) = 2x \qquad \text{[not quite what we wanted]}$$

We wanted the derivative to be x, but we got $2x$. This is easily fixed by halving our guess:

$$f(x) = \frac{1}{2}x^2 \implies f'(x) = x$$

The function $\frac{1}{2}x^2$ is an antiderivative of x. Therefore by the fundamental theorem we have

$$\begin{aligned}\int_0^1 x \ \mathrm{d}x &= f(1) - f(0) \\ &= \frac{1}{2}1^2 - \frac{1}{2}0^2 \\ &= \frac{1}{2}.\end{aligned}$$

This is the same result that we obtained earlier and with much more labor using Riemann sums.

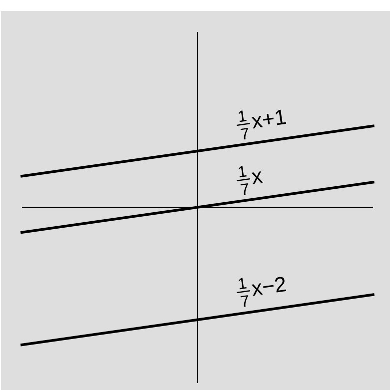
Because antiderivatives are so frequently used in order to evaluate definite integrals, expressions of the form $f(b) - f(a)$ are very common, and various abbreviations have been invented. We will abbreviate

$$f(b) - f(a) \stackrel{\text{def}}{=} f(x)\big]_{x=a}^{b} = f(x)\big]_a^b.$$

Any time we have an antiderivative, we can produce other antiderivatives by adding a constant. For example, all of the following are antiderivatives of the constant function $1/7$ with respect to x:

$$\frac{1}{7}x \qquad \frac{1}{7}x + 1 \qquad \frac{1}{7}x - 2$$

Differentiating any one of these with respect to x gives $1/7$.



k / All three functions are antiderivatives of the constant function 1/7. Shifting the graph vertically doesn't change its derivative.

Leibniz notation for the indefinite integral

An antiderivative is more commonly referred to as an indefinite integral — as opposed to the kind of integral we've been talking about up until now, which is called a definite integral. The Leibniz notation for an indefinite integral is an integral sign without any upper or lower limits of integration. For example,

$$\int x \ \mathrm{d}x = \frac{1}{2}x^2 + c,$$

where c is any constant. One way of understanding this notation is that both sides of this equals sign represent a certain *solution set* —

the set of all functions whose derivative equals x. Similarly, when we write

$$\sqrt{4} = \pm 2,$$

we could say that both sides of the equation represent the solution set $\{-2, 2\}$ of the equation $x^2 = 4$.

The following table summarizes the differences between definite and indefinite integrals.

indefinite integral	*definite integral*
$\int f(x)dx$ is a function of x.	$\int_a^b f(x)dx$ is a number.
By definition $\int f(x)\ \mathrm{d}x$ is any function of x whose derivative is $f(x)$.	$\int_a^b f(x)dx$ is defined in terms of Riemann sums and can be interpreted as the area under the graph of $y = f(x)$.
The variable of integration is not a dummy variable. For example, $\int 2x\ \mathrm{d}x = x^2 + c$ and $\int 2t\ \mathrm{d}t = t^2 + c$ are expressed in terms of different variables, so they are not the same.	The variable of integration is a dummy variable. For example, $\int_0^1 2x\ \mathrm{d}x = 1$, and $\int_0^1 2t\ \mathrm{d}t = 1$, so $\int_0^1 2x\ \mathrm{d}x = \int_0^1 2t\ \mathrm{d}t$.

Example 6

▷ Evaluate

$$\int x^6\ \mathrm{d}x$$

▷ Differentiation of a power will reduce the exponent by one, so we want something like x^7. The derivative of x^7 would be $7x^6$, which is too big by a factor of 7, so we want $x^7/7$. Including an arbitrary constant of integration, we have

$$\int x^6\ \mathrm{d}x = \frac{1}{7}x^7 + c.$$

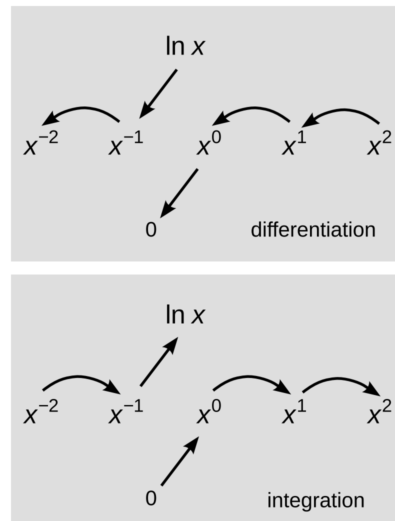
Integral of $1/x$ Example 7

▷ Evaluate the indefinite integral

$$\int \frac{\mathrm{d}x}{x}.$$

▷ As discussed in example 4, p. 179, this notation says that the function being integrated is $1/x$, or x^{-1}. Normally if we wanted to find the antiderivative of x to some power, we would increase the exponent by 1, as in example 6. But the derivative of x^0 is simply zero, so that doesn't work here. We recall that the ladder of powers is interrupted at this place, figure l. The indefinite integral we want is

$$\ln x + c.$$



l / Differentiation moves us down the ladder of powers of x. Integration climbs the ladder, as in example 6. Example 7 deals with the break in the middle of the ladder.

Area under the graph of $1/x$ Example 8

▷ Interpret the definite integral

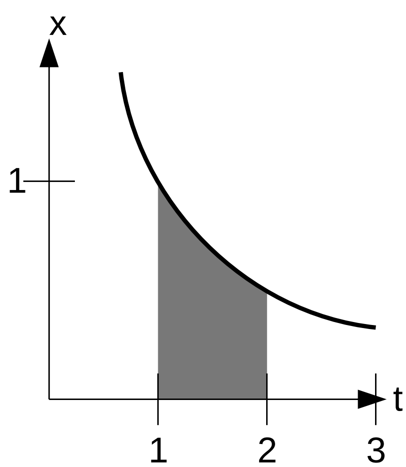
$$\int_1^2 \frac{dx}{x}.$$

graphically; then evaluate it .

▷ Figure m shows the graphical interpretation.

We saw in example 7 that the integral of $1/x$ was $\ln x + c$. Using the fundamental theorem of calculus, the area is $(\ln 2 + c) - (\ln 1 + c) \approx 0.693147180559945$. Note that the constant of integration cancels out when we plug in the upper and lower limits of integration and subtract; this always happens when we evaluate a definite integral in this way, so constants of integration are irrelevant in this context, and usually we would skip writing the $+c$.

Judging from the graph, it looks plausible that the shaded area is about 0.7.



m / Example 8.

8.4 Using the tool correctly

8.4.1 When do you need an integral?

In section 1.5.2, p. 23, we asked the question, "When do you need a derivative?" It's natural to ask the same question about integrals. And since the derivative and integral are so closely linked by the fundamental theorem of calculus, the answers should be related. If the relationship between two variables A and B is such that expressing A in terms of B requires a derivative, then expressing B in terms of A also requires calculus — it requires an integral.

As a concrete example, let x be your car's odometer reading, and let v be the reading on the speedometer. If v is constant, then we don't need calculus to express it in terms of x.

$$v = \frac{\Delta x}{\Delta t} \qquad \text{[only if } v \text{ is constant]} \tag{8}$$

But if v is changing, then equation (8) gives the wrong answer. We need calculus.

$$v = \frac{dx}{dt} \qquad \text{[always valid]} \tag{9}$$

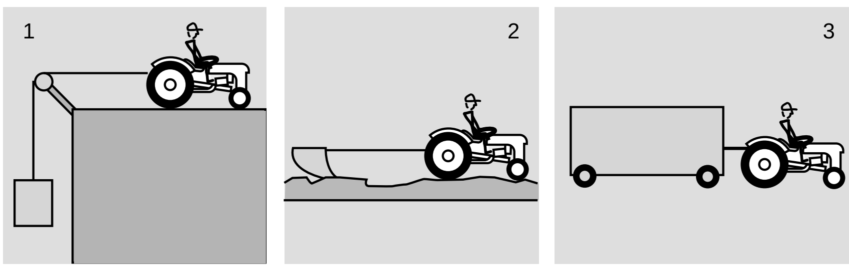
Now suppose we want x in terms of v. If v is constant, then we don't need calculus. Simple algebraic manipulation of equation (8) gives

$$\Delta x = v\,\Delta t. \qquad \text{[only if } v \text{ is constant]} \tag{10}$$

But equation (10) clearly doesn't make sense if v isn't constant. If you're in stop-and-go traffic, then your velocity isn't just one number. What would it even mean, then, to "multiply v by Δt?" Multiplication is like that special thing that happens when a mommy and a daddy love each other very much; it's something that happens between just one number and one other number. Applying the fundamental theorem of calculus to equation (9), we get

$$\Delta x = \int_{t_1}^{t_2} v \ \mathrm{d}t. \qquad \text{[always valid]} \tag{11}$$

We expect the integral to come up in applications as a generalization of multiplication that covers the case where one of the factors is varying.



n / Example 9. The tractor does mechanical work.

Work Example 9

▷ In each of the examples in figure n, the tractor exerts a force while traveling from position x_1 to position x_2, a distance $\Delta x = x_2 - x_1$. If the force F is constant, then the quantity $W = F\Delta x$, called mechanical work, measures the amount of energy expended. If W is the same in all three cases in the figure, then the amount of gas the tractor burns is identical in all three cases. How should this definition of mechanical work be generalized to the case where the force is varying?

▷ To generalize multiplication to a case where one of the factors isn't constant, we use an integral.

$$W = \int_{x_1}^{x_2} F \ \mathrm{d}x$$

8.4.2 Two trivial hangups

In section 1.4, p. 21, we discussed two common difficulties that students encounter in applying differentiation to real-world problems. The same two issues occur in integration. The first is that although a calculus textbooks will often notate every problem in terms of the letters y and x, any letters of the alphabet can occur in real-life applications. The second is that one often encounters symbolic constants, which are to be treated just like numerical constants.

A falling rock Example 10

▷ A falling rock has a velocity that increases linearly as a function of time, $v = at$, where a is a constant. Use an indefinite integral to find the position as a function of time.

▷ Let's first figure out the roles played by the three letters:

- t — the independent variable
- v — a function of t
- a — a constant
- x — the function we get as an indefinite integral

Next, let's warm up by translating this into a more stereotypical problem from a calculus textbook. For example, we could be given the function $y = 7x$ and asked to find its indefinite integral. The integral is $\int y\,\mathrm{d}x = (7/2)x^2 + c$.

The solution to the actual problem is found by simply shuffling letters of the alphabet and treating the constant a the same way we treated the constant 7. The setup of the integral is

$$x = \int v \, \mathrm{d}t,$$

and the result is

$$x = \frac{1}{2}at^2 + c.$$

The constant of integration is interpreted as the initial position, so it's actually nicer to give it a notation that indicates that:

$$x = \frac{1}{2}at^2 + x_o$$

8.4.3 Two ways of checking an integral

Every indefinite integral can be checked by taking its derivative to see if we can get back the original function. Furthermore, we can often check an integral by checking its units.

Checking the falling rock Example 11

Let's use these techniques to check the result of example 10. We were given the function

$$v = at. \tag{12}$$

We set up the integral as

$$x = \int v \, \mathrm{d}t, \tag{13}$$

and the result was

$$x = \frac{1}{2}at^2 + x_o. \tag{14}$$

First we take the derivative of both sides of equation (14). Because t is the independent variable here, these are derivatives with respect to t.

$$\frac{dx}{dt} \stackrel{?}{=} \frac{d}{dt}\left(\frac{1}{2}at^2 + x_o\right). \tag{15}$$

The left-hand side is the definition of the velocity v. On the right-hand side, we have to differentiate a polynomial. The constant a is treated like any other multiplicative constant: it just "comes along for the ride" in differentiation. The constant x_o is treated like any other additive constant in differentiation: it goes away.

$$v \stackrel{?}{=} a\frac{d}{dt}\left(\frac{1}{2}t^2\right) \tag{16}$$

The derivative of $(1/2)t^2$ with respect to t is t, so we recover equation 12, and our solution passes the check.

Next we check the units. The units of the given equation (12) ought to be right. If we remember the units of acceleration, we can check its units. If we don't remember the units of acceleration, we need to infer the units of the symbolic constant a from equation (12), because otherwise we won't be able to do the check on our own work. Based on equation (12), the units of acceleration are implied to be meters over seconds squared, $\mathrm{m/s^2}$.

Our initial setup in equation (13) has the following units:

$$\underbrace{x}_{\mathrm{m}} = \int \underbrace{v}_{\mathrm{m/s}} \underbrace{dt}_{\mathrm{s}}$$

The integral can be thought of as a sum, and the units of a sum are the same as the units of the things being added. This works out properly, so our setup passes this check as well.

We finish by checking the units of our final result, equation (14).

$$\underbrace{x}_{\mathrm{m}} = \underbrace{\frac{1}{2}}_{\text{unitless}} \underbrace{a}_{\mathrm{m/s^2}} \underbrace{t^2}_{\mathrm{s^2}} + \underbrace{x_o}_{\mathrm{m}}$$

8.4.4 Do I differentiate this, or do I integrate it?

In an end-of-chapter problem in a calculus textbook, you're usually commanded either to integrate or to differentiate. In real-world contexts, however, the question can arise of which one is the right thing to do. Often we have a pair of variables, and we know that one is the integral of the other, and one is the derivative of the other. But which one is which? Memorization would be the wrong way to approach this. The following is a list of possible ways of telling which is is which.

1. A derivative often represents a rate of change, an integral the accumulation of change.

2. Real-world quantities usually have units, and only one way of setting up the calculus relationship causes the units to make sense.

3. The integral often occurs as a generalization of multiplication, the derivative as a generalization of the slope of a line.

A chemical reaction Example 12

▷ Chemicals P and Q react to produce R. There is a reaction rate r and a concentration C of the product. Which would be the derivative of which, and which would be the integral of which?

▷ A derivative represents a rate of change, so $r = \mathrm{d}C/\mathrm{d}t$. An integral represents the accumulation of change, so $C = \int r\, \mathrm{d}t$.

An epidemic Example 13

▷ During an epidemic, there is some number of people I who have the disease, and some number w of new cases per day being reported. How would the calculus relationships between these two variables be set up?

▷ The variable I is unitless; it is just a *count* of the number of infected people. The variable w has units of cases per day, but "cases" is really a count, not a unit, so the units of w are really day^{-1} (inverse days). Conceptually, it's clear that these two quantities should be related as integral and derivative, and if we were unsure of which way around to write the relationship, the units would tell us.

$$\underbrace{w}_{\text{day}^{-1}} = \underbrace{\frac{\mathrm{d}I}{\mathrm{d}t}}_{\frac{\text{unitless}}{\text{days}}}$$

$$\underbrace{I}_{\text{unitless}} = \int \underbrace{w}_{\text{day}^{-1}} \underbrace{\mathrm{d}t}_{\text{days}}$$

An example of the third method was given in example 9, p. 187, where the definition of mechanical work was generalized to cases where the force varies.

8.5 Linearity

The most important and basic properties of the derivative (p. 16) are that it adds, $(f+g)' = f' + g'$, and scales vertically, $(cf)' = cf'$, where c is a constant. When an operation has these properties, we say that it is *linear*. Since the indefinite integral is defined as the antiderivative, it follows that the indefinite integral is also linear,

$$\int [f(x)+g(x)]\ \mathrm{d}x = \int f(x)\ \mathrm{d}x + \int g(x)\ \mathrm{d}x$$
$$\int cf(x)\ \mathrm{d}x = c\int f(x)\ \mathrm{d}x$$

and by the fundamental theorem the same is true for the definite integral.

Example 14

▷ Evaluate the definite integral

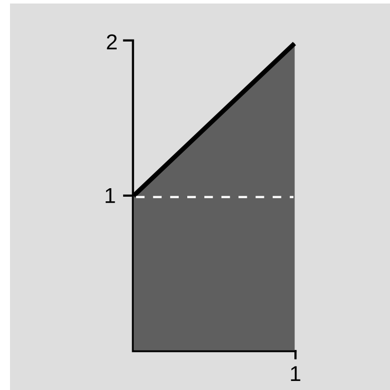
$$\int_0^1 (1+x)\ \mathrm{d}x$$

and give a geometrical interpretation.

▷ The linearity of the definite integral gives

$$\int_0^1 (1+x)\ \mathrm{d}x = \int_0^1 1\ \mathrm{d}x + \int_0^1 x\ \mathrm{d}x = 1 + \frac{1}{2} = \frac{3}{2}.$$

Figure o gives a geometrical interpretation.



o / Example 14. The total area is the area of the square base plus the area of the triangle on top.

8.6 Some technical points

8.6.1 Riemann sums in general

As a tree grows, its radius increases continuously. When a tree is cut down, as in figure p, we can see that the growth in each year is not the same. For example, in most of California, where the weather tends to be dry, a tree will usually show markedly increased growth in a wet year. In this example, it's natural to think of the radius of the tree as an integral of the form $\int_{\ldots}^{\ldots} \mathrm{d}r$. Of course it would be silly to try to explicitly calculate this integral, when we could simply measure the radius with a ruler! We don't really need calculus here, but, as is often the case, calculus guides us in thinking about the concepts even when we aren't going to use the techniques of calculus. If we were to approximate this integral using a Riemann sum, it would seem most natural to break the sum down into *unequal* intervals Δr. This is allowed by the definition of a Riemann sum, and the kind of Riemann sum that we defined on p. 175, with equal subintervals, was a more specific type.

p / Each tree ring adds Δr to the radius of the tree. The Δr values are not all the same.

A Riemann sum can also sample the value of the function at some other place than the center of each subinterval. The sample point can be at the left side, at the right, and it doesn't even need to be chosen in a consistent way for all the subintervals of a particular Riemann sum.

8.6.2 Integrating discontinuous functions

The definition of the integral given in section 8.2.1, for continuous functions, has some technical shortcomings if we try to apply it to badly behaved discontinuous functions. Most people who use calculus neither know nor care about these issues, and it's all right to skip this subsection on a first reading.

To show what can go wrong, we define two functions, one naughty and the other even naughtier.

- Let $f(x)$ be defined as $f(x) = 1/x$, except at $x = 0$, where we set $f(0) = 0$.

- Let $g(x)$ be the function such that if x is a rational number, $g(x) = 0$, but if x is irrational, $g(x) = 1$.

The definition of the integral in section 8.2.1 involved Riemann sums using equal subintervals, sampled at their centers. It carried a warning label saying that it only applied to continuous functions. Let's ignore the warning and see what goes wrong when we apply it to functions f and g.

The function f is discontinuous at only one point, and the discontinuity is one where it blows up to $+\infty$ on one side and $-\infty$ on the other. If we evaluate $\int_{-1}^{1} f(x)\ \mathrm{d}x$ using equal subintervals sampled at their centers, then because f is odd, every Riemann sum is exactly zero. The Riemann sums for odd n use $x = 0$ as a sample point, but these sums still vanish, because $f(0) = 0$. This integral, as defined in section 8.2.1, comes out to be zero.

The function g is what's known as a "pathological" example, meaning that it's so weird that we don't expect to encounter such a thing in any real-world application. For example, we could never determine a function like g from physical measurements, because measurements can't distinguish a rational number from an irrational one. If we evaluate $\int_0^1 g(x)\ \mathrm{d}x$ using equal subintervals, sampled at their centers, then every sample point is a rational number, so the integral comes out to be zero according to the definition in section 8.2.1.

The worrisome thing about both of these examples is that they both gave zero, but zero is either misleading or wrong in both cases. The result for the integral of f depended on a perfect cancellation of very large negative and very large positive terms in each Riemann sum. As n grew, these terms grew without bound, but they still canceled. In any real-world application, it's unlikely that this would happen. For example, if f represented the reading on a meter measuring the flow of water through a pipe (positive and negative indicating two different directions of flow), then the extreme positive and negative flows near $x = 0$ would have destroyed the meter!

The zero result for g is even more morally wrong. There are in some sense *more* irrational numbers than rational ones, so if this integral were to have some value, then clearly it should be 1, not 0.

What we would really like is to have our definition of the integral be stated in such a way that integrals like these come out to be undefined. This can be done by requiring in the definition that no matter what Riemann sum we use, regardless of whether the subintervals are equal or the sample points are at their centers, the limit must come out to be the same.

Definition of the integral (Riemann)
Suppose we have a number I such that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $|R - I| < \varepsilon$ for *every* Riemann sum all of whose intervals have width $x_{k+1} - x_k < \delta$, with any choice of sample points $s_1, \ldots, s_n$. Then I is the Riemann integral of the function.

For the integrals of the functions f and g described above, there is no number I with the properties described in the definition. The integral is then undefined, as it should be. A function for which such an I does exist is called Riemann integrable. A sufficient condition for Riemann integrability is that the function has only finitely many points of discontinuity, and it doesn't blow up at these discontinuities. For functions that are Riemann-integrable, the Riemann integral gives the same answer as the simpler definition in section 8.2, p. 175.

8.6.3 Proof of the fundamental theorem

We now refine the pseudo-proof in section 8.3.3, p. 182, into a real proof of the fundamental theorem of calculus. We want to prove that

$$\int_a^b f'(x)\,\mathrm{d}x = f(b) - f(a). \tag{17}$$

We assume that f' is Riemann integrable, so that we have the freedom to subdivide the interval $[a, b]$ and choose the sample points in any way that is convenient. We will break up the interval $[a, b]$ into n equal subintervals $[x_i, x_{i+1}]$, where $i = 1, 2, \ldots n - 1$. However, rather than restricting ourselves to sampling at the center of each subinterval, we apply the mean value theorem to each subinterval, and choose s_i to be the point for which

$$f'(s_i) = \frac{\Delta f_i}{\Delta x},$$

where $\Delta f_i = f(x_{i+1}) - f(x_i)$ and $\Delta x = x_{i+1} - x_i$. This can be rearranged to give

$$\Delta f_i = f'(s_i)\Delta x.$$

Adding these up, we have

$$f(b) - f(a) = \sum_{i=1}^{n} f'(s_i)\Delta x.$$

This tells us that by an appropriate choice of the sample points, we can make *every* Riemann sum, for *every* n produce the result claimed by the fundamental theorem. It therefore follows that the limit that defines the integral has the value claimed by the theorem.□

8.7 The definite integral as a function of its integration bounds

8.7.1 A function defined by an integral

Consider the expression

$$I = \int_0^x t^2\,dt.$$

What does I depend on? To find out, we calculate the integral

$$I = \left[\tfrac{1}{3}t^3\right]_0^x = \tfrac{1}{3}x^3 - \tfrac{1}{3}0^3 = \tfrac{1}{3}x^3.$$

So the integral depends on x. It does not depend on t, since t is a "dummy variable."

In this way we can use integrals to define new functions. For instance, we could define

$$I(x) = \int_0^x t^2\ \mathrm{d}t,$$

which would be a roundabout way of defining the function $I(x) = x^3/3$. Again, since t is a dummy variable we can replace it by any other variable we like. Thus

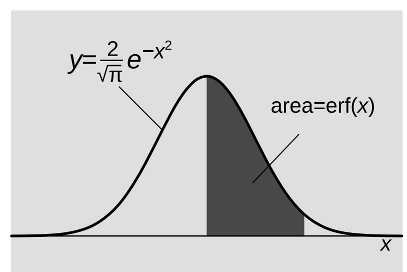
$$I(x) = \int_0^x \alpha^2\ \mathrm{d}\alpha$$

defines the same function (namely, $I(x) = \frac{1}{3}x^3$).

This example does not really define a new function, in the sense that we already had a much simpler way of defining the same function, by writing "$I(x) = x^3/3$." An example of a *new* function defined by an integral is the so called *error function* from statistics:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\ \mathrm{d}t, \tag{18}$$

so that $\mathrm{erf}(x)$ is the area of the shaded region in figure q.



q / The definition of the error function, erf(x).

The integral in (18) cannot be computed as a formula.[3] As described in more detail in section 10.1.2, p. 216, the integral in (18) occurs very often in statistics, so it has been given its own name, "$\mathrm{erf}(x)$".

8.7.2 How do you differentiate a function defined by an integral?

The answer is simple, for if $f(x) = F'(x)$ then the fundamental theorem says that

$$\int_a^x f(t)\ \mathrm{d}t = F(x) - F(a),$$

[3] For more on what this means, see section 9.3, p. 209.

and therefore

$$\frac{d}{dx}\int_a^x f(t)\ dt = \frac{\mathrm{d}}{\mathrm{d}x}\big(F(x) - F(a)\big) = F'(x) = f(x),$$

i.e.

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(t)\ \mathrm{d}t = f(x).$$

A similar calculation gives

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_x^b f(t)\ \mathrm{d}t = -f(x).$$

So what is the derivative of the error function? It is

$$\begin{aligned}\mathrm{erf}\,'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\ \mathrm{d}t\right]\\ &= \frac{2}{\sqrt{\pi}}\frac{\mathrm{d}}{\mathrm{d}x}\left[\int_0^x e^{-t^2}\ \mathrm{d}t\right]\\ &= \frac{2}{\sqrt{\pi}}e^{-x^2}.\end{aligned}$$

8.7.3 A second version of the fundamental theorem

The way that we differentiated the erf function in section 8.7.2 was an example of a more general idea, which can be considered as an alternative version of the fundamental theorem of calculus. The version of the fundamental theorem of calculus given in section 8.3, p. 181, says that if we differentiate and then integrate, we end up with the same function back again. This new second version says that something similar happens if we integrate and then differentiate:

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(t)\ \mathrm{d}t = f(x)$$

Problems

Problems a1-a3 don't require you to calculate anything. The point is to practice setting up and interpreting relationships between pairs of variables that are related as integral and derivative.

a1 A barometric altimeter is a device that uses a measurement of air pressure P to determine altitude y. Let the density of air be ρ (Greek letter "rho," the equivalent of Latin "r"), and the strength of the earth's gravitaty g. If ρ is constant, then the difference in pressure between two heights is given by

$$P_2 - P_1 = \rho g \Delta y.$$

Mountaineers and airplane pilots often traverse enough height that it is not a good approximation to take ρ as being constant; the air is less dense higher up. Use one of the methods of section 8.4.4, p. 189, to generalize the equation appropriately. ▷ Solution, p. 240

a2 Suppose that a business investment today will yield a stream of income in the future $f(t)$, in units of dollars per year. The revenue starts today, at $t = 0$, and will end in the future at $t = T$. The value of a dollar promised in the future is less than a dollar in hand today, because today's dollar could be put in the bank and draw interest, growing in value exponentially as e^{rt}, where r is a constant that is proportional to the interest rate. Consider the following two proposed expressions for the present value V of the revenue stream, i.e., the amount that one should rationally be willing to pay today in order to receive it.

$$\mathrm{V} = \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-rT} f(t)\right)$$

$$\mathrm{V} = \int_0^T f(t) e^{-rt}\ \mathrm{d}t$$

As described in section 8.4.4, p. 189, determine which of these is nonsense based on the units. ▷ Solution, p. 240

a3 An electric meter installed outside your household measures the flow of electric current I. If you turn on a lamp, I increases, and if you turn it back off again, I goes back down. The cost C of the electricity is also a function of time; it grows until it's time for the electric company to bill you. Consider the following two proposed relations between these variables.

$$I = k\frac{\mathrm{d}C}{\mathrm{d}t}$$

$$I = k\int_{t_1}^{t_2} C\ \mathrm{d}t$$

Here k is a constant. Use one of the methods of section 8.4.4, p. 189, to determine which of these makes sense. ▷ Solution, p. 241

c1 (a) Compute $\sum_{k=1}^{3} \frac{1}{k}$. (b) Compute $\sum_{m=1}^{3} \frac{1}{m}$. ✓

c2 (a) Which of the following are correct ways of notating the area of a right triangle with both legs of length 1?

$$\int_0^1 x \qquad \int_0^1 x \, \mathrm{d}x \qquad \int_0^1 u \, \mathrm{d}u$$

(b) The function f is defined by $f(x) = x^2 + 1$. Why is it wrong to notate the antiderivative of f as $\int x^2 + 1 \, \mathrm{d}x$? ▷ Solution, p. 241

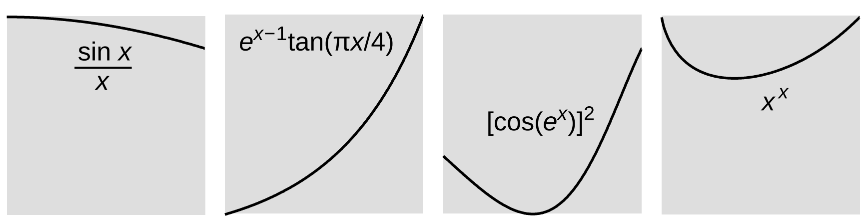
In each of problems c3-c6, the goal is to approximate the area between the graph and the x axis between $x = 0$ and $x = 1$, i.e., the value of $\int_0^1 f(x) \, \mathrm{d}x$ for the given function f. Each function was chosen such that for $x \in [0, 1]$, we have $y \in [0, 1]$ as well, so that the graph fits into a 1×1 square, as shown in the figure. These happen to be functions for which it is not possible to find an antiderivative, hence the need for an approximation. Divide the interval up into 5 equal subintervals, sample the function at the center of each interval, and find the resulting Riemann sum. Maintain four decimal places of precision throughout the calculation so that you are left with three decimal places at the end that are not likely to be way off simply because of rounding.

c3 $(\sin x)/x$ ✓

c4 $e^{x-1} \tan(\pi x/4)$ ✓

c5 $[\cos(e^x)]^2$ ✓

c6 x^x ✓



Problems c3-c6.

e1 Find three different functions of x whose derivatives with respect to x are all e^x. ▷ Solution, p. 241

e2 One or more of the following antiderivatives is incorrect. As described in section 8.4.3, use differentiation to find which are incorrect. Fix any incorrect ones.

$$\int x \, \mathrm{d}x = \frac{1}{2}x^2 + c \qquad \int e^{2x} \, \mathrm{d}x = e^{2x} + c$$

$$\int x^4 \, \mathrm{d}x = 4x^5 + c \qquad \int x^{-1} \, \mathrm{d}x = x^0 + c$$

$$\int e^x \, \mathrm{d}x = e^x + c$$

▷ Solution, p. 241

Evaluate the antiderivatives in problems e3-e14. If in doubt, guess and check as in problem e2. With experience it gets easier to guess correctly.

e3 $\displaystyle\int (2x+1)\ \mathrm{d}x$ ✓

e4 $\displaystyle\int (1-3t)\ \mathrm{d}t$ ✓

e5 $\displaystyle\int (u^2-u+11)\ \mathrm{d}u$ ✓

e6 $\displaystyle\int \left(1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}\right)\ \mathrm{d}x$ ✓

e7 $\displaystyle\int \frac{2\,\mathrm{d}q}{q}\qquad [q>0]$ ✓

e8 $\displaystyle\int \frac{e^a-e^{-a}}{2}\ \mathrm{d}a$ ✓

e9 $\displaystyle\int \frac{e^a+e^{-a}}{2}\ \mathrm{d}a$ ✓

e10 $\displaystyle\int \sin y\ \mathrm{d}y$ ✓

e11 $\displaystyle\int \cos y\ \mathrm{d}y$ ✓

e12 $\displaystyle\int \cos 2r\ \mathrm{d}r$ ✓

e13 $\displaystyle\int \sin(r-\pi/3)\ \mathrm{d}r$ ✓

e14 $\displaystyle\int (\sin x+\sin 2x)\ \mathrm{d}x$ ✓

Evaluate the antiderivatives in problems g1-g3. All letters other than the variable of integration are constants.

g1 $\displaystyle\int (Ax+B)\ \mathrm{d}x$ ✓

g2 $\displaystyle\int bx^a\ \mathrm{d}x\qquad [a\neq -1]$ ✓

g3 $\displaystyle\int \cos\omega\tau\ \mathrm{d}\tau$ ✓

g4 $\displaystyle\int e^{\beta t}\ \mathrm{d}t$ ✓

In problems i1-i4, find the antiderivatives. All letters other than the variable of integration are constants. These problems can be done by first rewriting the given integrand in a form that you know how to integrate.

i1 $\int \sqrt{Bx\sqrt{x}}\,\mathrm{d}x$ ▷ Solution, p. 242

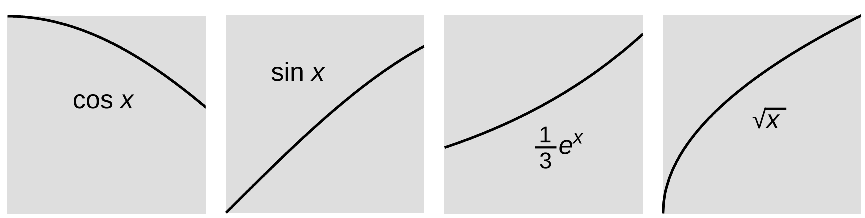
i2 $\int e^{z+\beta}\,\mathrm{d}z$ √

i3 $\int \frac{p^2+k}{\sqrt{p}}\,\mathrm{d}p$ √

i4 $\int \frac{pw-q}{\sqrt[3]{w^2}}\,\mathrm{d}w$ √

These instructions are for problems k1-k4. Each function f was chosen such that for $x \in [0,1]$, we have $y \in [0,1]$ as well, so that the graph fits into a 1×1 square, as shown in the figure.
(a) Make an eyeball estimate of the area under the curve.
(b) As in problems c3-c6, divide the interval up into 5 equal subintervals, sample the function at the center of each interval, and find the resulting Riemann sum. Maintain four decimal places of precision throughout the calculation so that you are left with three decimal places at the end that are not likely to be way off simply because of rounding. Your result should be roughly consistent with your estimate from part a, and you can also check it online.
(c) Find the antiderivative $\int f(x)\,\mathrm{d}x$, and check it online.
(d) Evaluate the definite integral, $\int_0^1 f(x)\,\mathrm{d}x$, check it against the approximations in parts a and b, and check it online.

Problems k1-k4.



k1 $f(x) = \cos x$ √

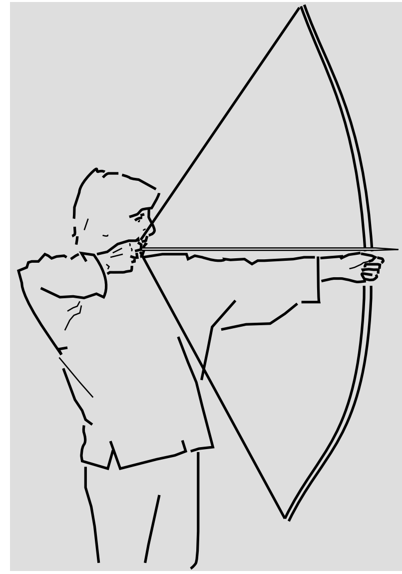
k2 $f(x) = \sin x$ √

k3 $f(x) = \frac{1}{3}e^x$ √

k4 $f(x) = \sqrt{x}$ √

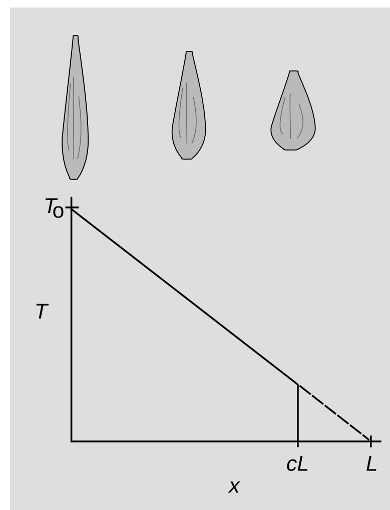
Problems n1-n4 all involve calculating the work done by a force, as described in example 9, p. 187. These problems also require you to check the units of your result. To do that, you will need to know the following. The SI unit of force is the newton (N). Work has units of (*force*) $\times$ (*distance*)*, or* N·m *(newton-meters).*

n1 The figure shows an archer drawing a longbow. When the string is pulled back to a distance x relative to its straight equilibrium position, the force required from the right hand is given approximately by $F = kx$, where k is a constant. (a) Infer the units of k. (b) Find the amount of work done in pulling the bow from $x = 0$ to $x = b$, where b is some number. (c) Check that the units of your result make sense. ▷ Solution, p. 242



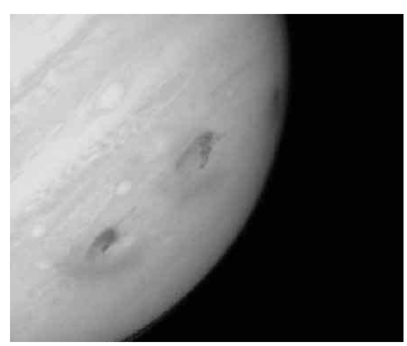
Problem n1.

n2 The figure shows the tension (force) of which a muscle is capable. The variable x is defined as the contraction of the muscle from its maximum length L, so that at $x = 0$ the muscle has length L, and at $x = L$ the muscle would theoretically have zero length. In reality, the muscle can only contract to $x = cL$, where c is less than 1. When the muscle is extended to its maximum length, at $x = 0$, it is capable of the greatest tension, T_o. As the muscle contracts, however, it becomes weaker. There is a nearly linear decrease, which would theoretically extrapolate to zero at $x = L$. (a) Infer the units of c and T_o. (b) Find the maximum work the muscle can do in one contraction, in terms of c, L, and T_o. (c) Show that your answer to part b has the right units. (d) Show that your answer to part b has the right behavior when $c = 0$ and when $c = 1$. ✓

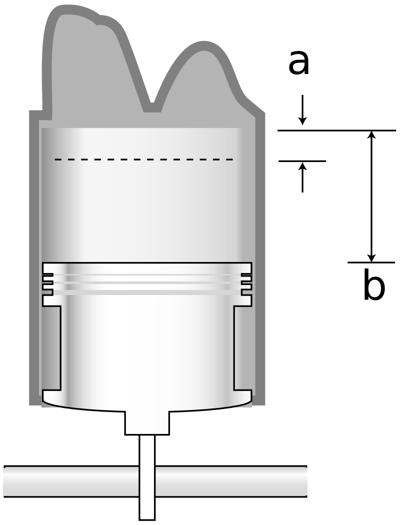


Problem n2.

n3 In July 1994, Comet Shoemaker-Levy 9, which had previously broken up into pieces, collided with the planet Jupiter. The figure shows discolorations left in the jovian atmosphere where the impacts had occurred. The diameter of each bruise is on the same order of magnitude as the size of the planet earth. These were hard hits. The energy came from the work done by the sun's gravity on the comet as it fell inward from the Oort Cloud, a hypothesized outer region of the solar system. Let x be the comet's position relative to the sun, and assume that the comet falls in from the negative x direction, i.e., from the side of the sun that we would visualize as the left-hand side of the number line. The force of the sun's gravity on the comet is given by Newton's law of gravity, $F = GMm/x^2$, where M is the mass of the sun, m is the mass of the comet, G is a universal constant, and the plus sign indicates that the force is to the right, i.e., toward the sun.
(a) Infer the units of G. (b) Find the work done on the comet as it falls from $x = -a$ to $x = -b$, where a is the distance from the sun to the Oort cloud, b is the distance from the sun to Jupiter, and both a and b are positive. (c) Check that the units of your answer to part b make sense. ✓



Problem n3.



Problem n4.

n4 *See the instructions on p. 201.* In a gasoline-burning car engine, the exploding air-gas mixture makes a force on the piston, and the force tapers off as the piston expands, allowing the gas to expand. A not-so-bad approximation is that the force is given by $F = k/x$, where x is the position of the piston. (a) Infer the units of k. (b) Find the work done on the piston as it travels from $x = a$ to $x = b$. (c) Show that the result of part b can be reexpressed so that it depends only on the ratio b/a. This ratio is known as the compression ratio of the engine. (d) Check that the units of your result in part c make sense. ✓

q1 If a car on cruise control has the wrong speed at $t = 0$, it will take some time for the system to correct the error. The system may be designed to produce a velocity as a function of time given by

$$v = u + be^{-rt},$$

where u is the desired speed, r is a constant chosen by the designer, and b is the initial error in velocity, which may be positive or negative. The value of r is a design compromise; if r is too small, then it will take a long time for the car to get back to the right speed, but if it is too big, the motion will be jerky or produce bad fuel efficiency.
(a) Infer the units of u, b, and r.
(b) Find the position x as a function of time. ✓
(c) Give a physical interpretation of the constant of integration occurring in your answer to part b.
(d) Check that your answer to part b has units that make sense.
(e) Check your answer by differentiating it.

q2 A piston in a car's engine is connected to the crankshaft through a piston rod. As the crankshaft spins at a constant rate, the velocity of the piston in and out of the cylinder may be approximated by a function

$$v = A\cos\omega t + B\cos 2\omega t,$$

where ω (Greek letter "omega," which makes the "o" sound) is the number of radians per second at which the crankshaft is rotating, and A and B are constants that depend on the length of the piston rod and the radius of the circle traveled by the piston pin. Note that expressions of the form $\sin xy$ are normally to be read as $\sin(xy)$; if the intended meaning had been $(\sin x)y$, then one would normally have written it as $y\sin x$.
(a) Infer the units of A and B. (The units of ω are simply inverse seconds, s^{-1}.)
(b) Find the piston's position x as a function of time. ✓
(c) Give a physical interpretation of the constant of integration occurring in your answer to part b.
(d) Check that your answer to part b has units that make sense.
(e) Check your answer by differentiating it.

In problems s1-s12, compute the definite integrals. These are in groups of three similar problems, with the intention being that a given student would do one from each group.

s1 $\displaystyle\int_1^2 u^{-2}\ \mathrm{d}u$ $\checkmark$

s2 $\displaystyle\int_1^2 w^{-3}\ \mathrm{d}w$ $\checkmark$

s3 $\displaystyle\int_1^2 s^{-1/2}\ \mathrm{d}s$ $\checkmark$

s4 $\displaystyle\int_0^1 (2h^3 - 3h + 1)\ \mathrm{d}h$ $\checkmark$

s5 $\displaystyle\int_0^1 (z^2 + 7z)\ \mathrm{d}z$ $\checkmark$

s6 $\displaystyle\int_0^1 (2r^4 - 2r^2 + r)\ \mathrm{d}r$ $\checkmark$

s7 $\displaystyle\int_0^4 (e^{2g} + \sin g - \sqrt{g})\ \mathrm{d}g$ $\checkmark$

s8 $\displaystyle\int_1^4 \left(\frac{1}{a} - a^{-3/2} + \cos a\right)\ \mathrm{d}a$ $\checkmark$

s9 $\displaystyle\int_0^4 (\cos p + e^{-p} + p^3)\ \mathrm{d}p$ $\checkmark$

s10 $\displaystyle\int_0^1 u(\sqrt{u} + \sqrt[3]{u})\ \mathrm{d}u$ $\checkmark$

s11 $\displaystyle\int_{-1}^1 (x-1)(3x+2)\ \mathrm{d}x$ $\checkmark$

s12 $\displaystyle\int_1^2 \left(j + \frac{1}{j}\right)^2\ \mathrm{d}j$ $\checkmark$

u1 Is the following calculation wrong? Explain why or why not.

$$\int_0^1 x\ \mathrm{d}x = \frac{1}{2}x^2 + 42\Big]_0^1 = \frac{1}{2}$$

u2 Let the functions f and g be defined as follows.

$$f(x) = \begin{cases} \ln(-x) + 7 & \text{if } x < 0 \\ \ln x + 11 & \text{if } x > 0 \end{cases}$$

$$g(x) = \ln|x|$$

Is f an antiderivative of $1/x$? Is g? Explain why or why not.

Chapter 9

Basic techniques of integration

9.1 Doing integrals symbolically on a computer

The quaint little town of Carmel, California, has a touristy business district that specializes in quaint little shops. I once went into a yarn store there with my mother, who picked out two skeins of yarn for a sweater. The business ran on paper and pen, which was arguably sensible, since there was little room on the cramped counter for a cash register. The following math problem resulted:

$$\begin{array}{r} \$5.60 \\ \times \quad 2 \end{array}$$

The proprietor pulled out a calculator and typed $0 \times 2 =$. The answer was 0, which she wrote down. Then $6 \times 2 =$, and so on.

The point of this anecdote is that there are right ways and wrong ways to use tools. Computers are a good tool for doing integrals, but we should be able to do simple integrals by hand.

The computer programs used for doing integrals are called computer algebra systems (CAS). I recommend a free and open-source CAS called Maxima.[1] The following example shows how to use Maxima to do an easy indefinite integral — analogous to using the calculator to find 6×2. The typewriter font shows what I typed in, and the italicized text is the answer printed out by the program. Note the mandatory semicolon at the end of the input line.

Integrating on a computer Example 1

```
integrate(cos(x),x);
```

sin(x)

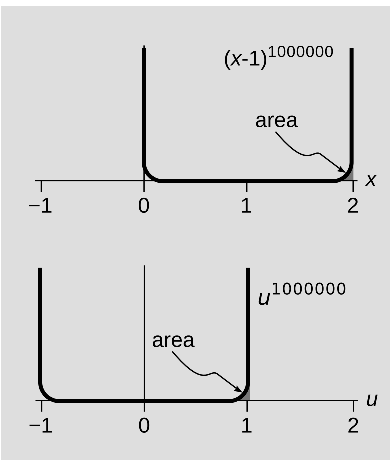
[1]To use it through a web browser go to `maxima-online.org`. To download it to your computer, go to `maxima.sourceforge.net`.

9.2 Substitution

Here's an example of an integral that introduces a useful technique of integration, and that also demonstrates what can go wrong if you become completely dependent on computers to do integrals that you should be able to do by hand.

$$\int_1^2 (x-1)^{1000000}\ \mathrm{d}x$$

I tested this on three CAS programs, and although two were able to do it, one froze up indefinitely. My point is not that a certain CAS is better than some other one.[2] The point is that computers, unlike humans, can't step back and say, "Hey, what I'm doing isn't working so well. Maybe I should try something else." The one that failed presumably started grinding away to multiply out the polynomial — all million and one terms of it: $x^{1000000} - 1000000x^{999999} + \ldots$ This is certainly a strategy that would work, in theory, because it would reduce the problem to one that we already know how to solve: integrating a polynomial.

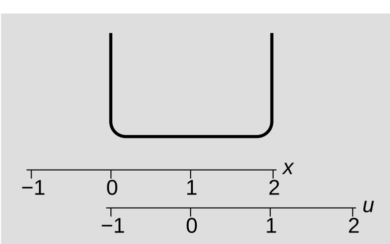


a / Integrating $(x-1)^{1000000}$ by using a change of variable. The function is not drawn realistically; the rounded edge has been exaggerated in order to make the shaded area under the curve visible.

But there's a better way to approach this, as suggested in figure a. Geometrically, what we're trying to calculate is the very small area that is only visible at the corner of the figure. (Although the limits of integration run from 1 to 2, the value of the integrand is too small to matter except when x gets very close to 2.) Let's shift the graph to the left by one unit, as shown in the figure, and define a new variable $u = x - 1$. The shift to the left doesn't change the amount of area under the curve; it simply *relocates* that area to a new place. In terms of this variable, the integrand is $u^{1000000}$, which is a function that we know how to integrate. Expressed as an integral with respect to u, the limits of integration are from $u = 1 - 1 = 0$ to $u = 2 - 1 = 1$. Do we need to do anything to the $\mathrm{d}x$ other than change it to a $\mathrm{d}u$? Not in this case; implicit differentiation of $u = x - 1$ gives $\mathrm{d}u = \mathrm{d}x$. The result is that we can calculate the same area using the following easier integral.

$$\int_0^1 u^{1000000}\ \mathrm{d}u$$

This is easily found to equal 1/1000001.



b / The change of variables just renames the points on the horizontal axis.

Figure b shows a nice way of thinking about this. Rather than imagining that the graph itself has shifted horizontally, we can say that the graph stayed in the same place, but we slid the axis over. This is just like renaming the points on the horizontal axis. The renaming is like sliding a ruler over without shrinking or expanding the ruler. If we think of $\mathrm{d}x$ as a small *change* in x, and similarly for $\mathrm{d}u$, then it makes sense that $\mathrm{d}u = \mathrm{d}x$; the distance or *difference*

[2] For the record, the two that could handle it were Maxima and `integrals.com`. The one that failed was another open-source program called Yacas.

between two points on a ruler is the same regardless of whether we slide the ruler around.

The procedure demonstrated above is called a change of variable, substitution, or sometimes "u-substitution," since it seems to be common for calculus textbooks to use the letter u in this context. In general, u can be defined as any function of x that you think will help to massage the integral into a more workable form. Substitution can be used on both definite and indefinite integrals.

Substitution with rescaling — Example 2

A common rate of return on ultra-safe, ten-year bonds has historically been about 5%, which means that money invested in these bonds grows by a factor of e in about $1/\ln 1.05 \approx 20$ years. Therefore we expect such an investment to grow exponentially over time in proportion to the function $e^{t/20}$, where t is in years. Bonds often pay dividends, and although the dividend payments actually occur at discrete time intervals, it can be convenient to model them mathematically as if they were paid continuously, so that the total dividend payment is

$$D = k \int_0^{10} e^{t/20}\ \mathrm{d}t,$$

where k is a constant. Let's evaluate this integral.

Since the derivative of e^x is e^x, we know how to integrate e^x, and it's natural to look for a substitution that makes the integrand into this form. The substitution clearly has to be

$$u = \frac{t}{20}. \tag{1}$$

If we think of the time axis as a "time-line" like the ones in history books, then this substitution is like expanding the time-line's scale by a factor of 20. Solving for $t = 20u$ and applying implicit differentiation gives

$$\mathrm{d}t = 20\ \mathrm{d}u. \tag{2}$$

The limits of integration change when expressed in terms of u.

$$t = 0 \qquad \Leftrightarrow \qquad u = 0 \tag{3}$$

$$t = 10 \qquad \Leftrightarrow \qquad u = \frac{1}{2} \tag{4}$$

We have to make use of all four of the equations (1)-(4) in order to rewrite the integral in terms of the new variable u:

$$\begin{aligned} D &= k \int_0^{1/2} e^u\,(20\,\mathrm{d}u) \\ &= 20k\ e^u\Big]_0^{1/2} \\ &= 20k\left(e^{1/2} - 1\right) \end{aligned}$$

A nonlinear substitution Example 3

▷ Evaluate

$$\int 2x\sin(x^2+3)\ \mathrm{d}x.$$

▷ Here the only substitution that has any hope of working is $u = x^2+3$. Implicit differentiation gives $\mathrm{d}u = 2x\,\mathrm{d}x$, which happens to be exactly the combination of factors that occurs in the integrand. The integral therefore equals:

$$\begin{aligned}\int \sin u\ \mathrm{d}u &= -\cos u + c\\ &= -\cos(x^2+3)+c\end{aligned}$$

To check that this indefinite integral is correct, we can differentiate it, which involves using the chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(-\cos(x^2+3)+c\right) = \sin(x^2+3)\cdot 2x$$

The method used to check example 3 shows that we should be able to interpret what's going on in these substitutions in terms of the chain rule. The chain rule says that

$$\frac{\mathrm{d}F(G(x))}{\mathrm{d}x} = F'(G(x))\cdot G'(x),$$

so that

$$\int F'(G(x))\cdot G'(x)\ \mathrm{d}x = F(G(x))+c.$$

In example 3, we had $2x = \frac{\mathrm{d}}{\mathrm{d}x}(x^2+3)$. So let's call $G(x)=x^2+3$, and $F(u)=-\cos u$. Then

$$F(G(x)) = -\cos(x^2+3)$$

and

$$\frac{\mathrm{d}F(G(x))}{\mathrm{d}x} = \underbrace{\sin(x^2+3)}_{F'(G(x))}\cdot\underbrace{2x}_{G'(x)} = f(x),$$

so that

$$\int 2x\sin(x^2+3)\ \mathrm{d}x = -\cos(x^2+3)+c.$$

9.3 Integrals that can't be done in closed form

Integral calculus was invented in the age of powdered wigs and harpsichords, so the original emphasis was on expressing integrals in a form that would allow numbers to be plugged in for easy numerical evaluation by scribbling on scraps of parchment with a quill pen. This was an era when you might have to travel to a large city to get access to a table of logarithms.

In this computationally impoverished environment, one always wanted to get answers in what's known as *closed form* and in terms of *elementary functions*.

A closed form expression means one written using a finite number of operations, as opposed to something like the geometric series $1 + x + x^2 + x^3 + \ldots$, which goes on forever.

Elementary functions are usually taken to be addition, subtraction, multiplication, division, logs, and exponentials, as well as other functions derivable from these. For example, a cube root is allowed, since $\sqrt[3]{x} = e^{(1/3)\ln x}$, and so are trig functions and their inverses, because they can be expressed in terms of logs and exponentials by using Euler's formula.

In theory, "closed form" doesn't mean anything unless we state the elementary functions that are allowed. In practice, when people refer to closed form, they usually have in mind the particular set of elementary functions described above.

A traditional freshman calculus course spends such a large amount of time teaching you how to do integrals in closed form that it may be easy to miss the fact that this is impossible for the vast majority of integrands that you might randomly write down. Here are some examples of impossible integrals:

$$\int e^{-x^2}\,\mathrm{d}x$$
$$\int x^x\,\mathrm{d}x$$
$$\int \frac{\sin x}{x}\,\mathrm{d}x$$
$$\int e^x \tan x\,\mathrm{d}x$$

The first of these is a form that is extremely important in statistics (it describes the area under the standard "bell curve"), so you can see that impossible integrals aren't just obscure things that don't pop up in real life.

People who are proficient at doing integrals in closed form generally seem to work by a process of pattern matching. They recognize certain integrals as being of a form that can't be done, so they know not to try.

Disobedience Example 4

▷ Students! Stand at attention! You will now evaluate $\int e^{-x^2+7x}\,dx$ in closed form.

▷ No sir, I can't do that. By a change of variables of the form $u = x + c$, where c is a constant, we could clearly put this into the form $\int e^{-x^2}\,dx$, which we know is impossible.

Sometimes an integral such as $\int e^{-x^2}\,dx$ is important enough that we want to give it a name, tabulate it, and write computer subroutines that can evaluate it numerically. For example, statisticians define the "error function" $\operatorname{erf}(x) = (2/\sqrt{\pi})\int e^{-x^2}dx$. Sometimes if you're not sure whether an integral can be done in closed form, you can put it into computer software, which will tell you that it reduces to one of these functions. You then know that it can't be done in closed form. For example, if you ask `integrals.com` to do $\int e^{-x^2+7x}\,dx$, it spits back $(1/2)e^{49/4}\sqrt{\pi}\operatorname{erf}(x-7/2)$. This tells you both that you shouldn't be wasting your time trying to do the integral in closed form and that if you need to evaluate it numerically, you can do that using the erf function.

As shown in the following example, just because an indefinite integral can't be done, that doesn't mean that we can never do a related definite integral.

▷ Example 5

Evaluate $\int_0^{\pi/2} e^{-\tan^2 x}(\tan^2 x + 1)\,dx$.

▷ The obvious substitution to try is $u = \tan x$, and this reduces the integrand to e^{-x^2}. This proves that the corresponding indefinite integral is impossible to express in closed form. However, the definite integral *can* be expressed in closed form; it turns out to be $\sqrt{\pi}/2$.

Sometimes computer software can't say anything about a particular integral at all. That doesn't mean that the integral can't be done. Computers are stupid, and they may try brute-force techniques that fail because the computer runs out of memory or CPU time. For example, the integral $\int dx/(x^{10000} - 1)$ can be done in closed form, and it's not too hard for a proficient human to figure out how to attack it, but every computer program I've tried it on has failed silently.

9.4 Doing an integral using symmetry or geometry

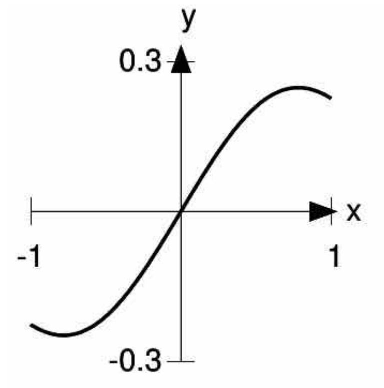
Often we can figure out the value of an integral either by symmetry or by using simple geometry.

An integral that vanishes by symmetry Example 6

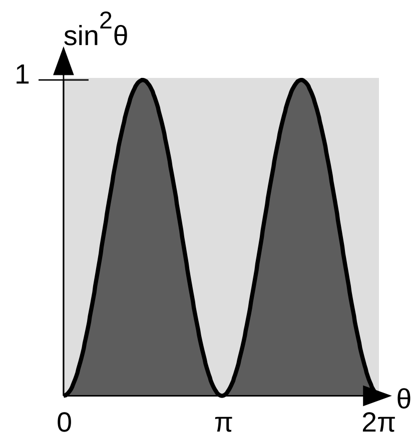
▷ Evaluate

$$\int_{-1}^{1} \frac{\sin x \, dx}{1+e^{x^2}}.$$

▷ I doubt that this can be done by finding the indefinite integral and plugging in the limits of integration. I tried it using the open-source program Maxima, and also using the web interface to a proprietary program called Mathematica, and neither could do it. However, the function is odd because the numerator is odd and the denominator is even. Since the function is odd, and the limits of integration are symmetrically placed on either side of the origin, the definite integral is guaranteed to be zero; any negative contribution to the integral on the left is guaranteed to be canceled by a matching positive contribution on the right.



c / The integrand of example 6.



d / The integrand of example 7.

An integral that can be done by geometry Example 7

▷ Evaluate

$$\int_{0}^{2\pi} \sin^2\theta \, d\theta.$$

▷ The hard way to do this integral is to dig up the appropriate trig identity, which allows $\sin^2\theta$ to be reexpressed in terms of $\sin 2\theta$. The easy way is to look at the graph, figure d. The rectangle is exactly half filled by the area under the graph. Since the rectangle has area 2π, the integral equals π.

9.5 Some forms involving exponentials, rational functions, and roots

Here are some forms whose antiderivatives may not be obvious at first sight.

9.5.1 Exponentials with the base not e

Since the derivative of e^x with respect to x is just e^x again, we already know how to integrate e^x. What about exponentials with other bases? These can be converted into base e using the identity $a^b = e^{b\ln a}$, then integrated using a change of variable.

Example 8

▷ Evaluate $\int 3^x \, dx$.

▷

$$\begin{aligned}
\int 3^x \, dx &= \int e^{x\ln 3} \, dx && [\text{using } a^b = e^{b\ln a}] \\
&= \frac{1}{\ln 3}\int e^u \, du && [\text{substituting } u = x\ln 3] \\
&= \frac{e^u}{\ln 3} \\
&= \frac{e^{x\ln 3}}{\ln 3} \\
&= \frac{3^x}{\ln 3} && [a^b = e^{b\ln a} \text{ again}]
\end{aligned}$$

9.5.2 Some forms involving rational functions and roots

In sections 5.10-5.11, pp. 137-137, we summarized the derivatives of various transcendental functions. Each of these potentially gives some way to integrate something, by applying the fundamental theorem. Some of these derivatives are not themselves transcendental functions, which makes it not at all obvious when looking at them that they should be attacked in this way:

derivative	*integral*
$(\tan^{-1} x)' = (1+x^2)^{-1}$	$\int (1+x^2)^{-1} \, dx = \tan^{-1} x + c$
$(\tanh^{-1} x)' = (1-x^2)^{-1}$	$\int (1-x^2)^{-1} \, dx = \tanh^{-1} x + c$
$(\sin^{-1} x)' = (1-x^2)^{-1/2}$	$\int (1-x^2)^{-1/2} \, dx = \sin^{-1} x + c$
$(\sinh^{-1} x)' = (x^2+1)^{-1/2}$	$\int (x^2+1)^{-1/2} \, dx = \sinh^{-1} x + c$
$(\cosh^{-1} x)' = (x^2-1)^{-1/2}$	$\int (x^2-1)^{-1/2} \, dx = \cosh^{-1} x + c$

Problems

In problems a1-a12, evaluate the indefinite integrals. Check your answer by differentiating it, and also check it online. All letters other than the variable of integration are constants. These are in groups of three similar problems, with the intention being that a given student would do one from each group.

a1 $\int \frac{\mathrm{d}f}{2f-4}$ $[f>2]$ ✓

a2 $\int \frac{\mathrm{d}w}{1-w}$ $[w<1]$ ✓

a3 $\int \frac{\mathrm{d}q}{q}$ $[q<0]$ ✓

a4 $\int 2^{cx}\ \mathrm{d}x$ ✓

a5 $\int c^s\ \mathrm{d}s$ $[c>0]$ ✓

a6 $\int 10^{a+\delta}\ \mathrm{d}\delta$ ✓

a7 $\int \frac{\mathrm{d}t}{a^2+t^2}$ ✓

a8 $\int \frac{\mathrm{d}v}{\left(\frac{v}{k}\right)^2+1}$ ✓

a9 $\int \frac{\mathrm{d}\phi}{\sqrt{A^2-\phi^2}}$ $[A>0]$ ✓

a10 $\int \cos^n \zeta \sin \zeta\ \mathrm{d}\zeta$
[$n \neq -1$; ζ is lowercase Greek zeta, which makes the "z" sound.] ✓

a11 $\int e^{e^\lambda} e^\lambda\ \mathrm{d}\lambda$
(λ is lowercase Greek lambda, which makes the "l" sound.) ✓

a12 $\int e^{\sin p} \cos p\ \mathrm{d}p$ ✓

In c1-c6, use a substitution to evaluate the indefinite integrals.

c1 $\int \sin\left(\frac{\pi+x}{5}\right) \mathrm{d}x$ ✓

c2 $\int \frac{\sin 2x \ \mathrm{d}x}{\sqrt{1+\cos 2x}}$ ✓

c3 $\int \frac{\sin 2x \ \mathrm{d}x}{1+\cos^2 x}$ ✓

c4 $\int \frac{\sin 2x \ \mathrm{d}x}{1+\sin x}$ ✓

c5 $\int \alpha e^{-\alpha^2} \ \mathrm{d}\alpha$ ✓

c6 $\int \frac{e^{1/t}}{t^2} \ \mathrm{d}t$ ✓

c7 $\int (z+3)\sqrt{z-1}\,\mathrm{d}z$ ✓

In e1-e9, use a substitution to evaluate the definite integrals.

e1 $\int_1^2 \frac{u \ \mathrm{d}u}{1+u^2}$ ✓

e2 $\int_1^2 \frac{\mu^2 \ \mathrm{d}\mu}{\mu^3+1}$ ✓

e3 $\int_0^5 \frac{x \ \mathrm{d}x}{\sqrt{x+1}}$ ✓

e4 $\int_1^2 \frac{x^2 \ \mathrm{d}x}{\sqrt{2x+1}}$ ✓

e5 $\int_0^\pi \cos\left(\theta+\pi/3\right) \mathrm{d}\theta$ ✓

e6 $\int_{\pi/4}^{\pi/3} \sin^3\theta\cos\theta \ \mathrm{d}\theta$ ✓

e7 $\int_0^{\sqrt{2}} \xi(1+2\xi^2)^{10} \ \mathrm{d}\xi$ ✓

e8 $\int_2^3 \frac{\mathrm{d}r}{r\ln r}$ ✓

e9 $\int_1^2 \frac{\ln 2x}{x} \ \mathrm{d}x$ ✓

In problems g1-g2, two indefinite integrals are given that involve functions which look similar to one of the following:

$$e^{-x^2} \qquad x^x \qquad \frac{\sin x}{x} \qquad e^x \tan x$$

As discussed in section 9.3, the four functions given above can't be integrated in closed form. In each pair below, one can be integrated, while the other can be made into one of the above forms by a substitution, proving that it's impossible to integrate. Determine which is which, integrate the one that can be done, and check your answer to that one online.

g1 (a) $\int x^{-3/4}e^{-\sqrt{x}} \ \mathrm{d}x$ (b) $\int x^{-1/2}e^{-\sqrt{x}} \ \mathrm{d}x$ ✓

g2 (a) $\int x^{-2}\sin\frac{1}{x} \ \mathrm{d}x$ (b) $\int x^{-1}\sin\frac{1}{x} \ \mathrm{d}x$ ✓

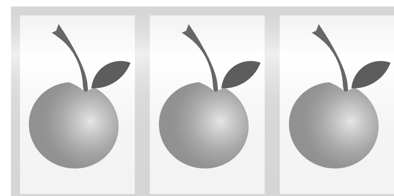
Chapter 10
Applications of the integral

10.1 Probability

10.1.1 Introduction to probability

Measurement of probabilities

Defining randomness is a difficult problem, tied up with classical philosophical issues such as determinism and free will. Mathematicians sidestep this question by simply using numbers between 0 and 1 to represent probabilities. A zero probability represents an event that can't happen, a probability of 1 an event than is guaranteed to happen. In between we have things that might or might not happen. A flipped coin comes up heads with probability 1/2.



a / The probability that one wheel on the slot machine will give a cherry is 1/10. If the three probabilities are independent, then the probability that all three wheels will give cherries is $1/10 \times 1/10 \times 1/10$.

Statistical independence

When ordinary people say that an event is "random," they usually mean not just that it has a probability greater than 0 and less than 1, but also that it can't be predicted, because there is no way of finding a connection with another event that caused it. This lack of connection is considered by mathematicians to be separate from randomness itself, and is defined as follows.

> **Definition of statistical independence**
> Events A and B are said to be statistically independent if the probability that they will both happen is given by the product of the two probabilities.

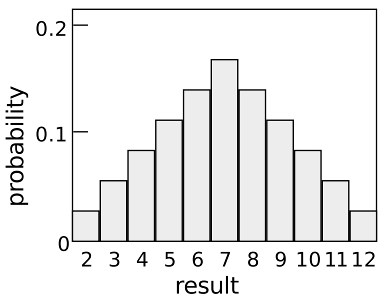
Events can be random but not independent. It might or might not rain tomorrow, and there might or might not be a forest fire. These events are both random, but they are not independent, since rain makes fire less likely.

b / The earth's surface is 30% land and 70% water. If we spin a globe and pick a random point, the probabilities of hitting land and water are 0.3 and 0.7. Normalization requires that these two probabilities add up to 1.

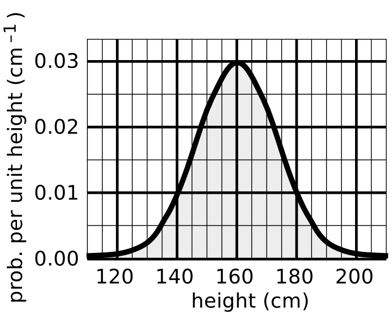
Normalization

Suppose that we are able to exhaustively list all of the possible outcomes A, B, C, ... of some situation, and that these outcomes are mutually exclusive. Then exactly one of these outcomes must occur, so the probabilities must add up to one. For example, suppose that we flip a coin, and A is the event that the coin comes up heads, B tails. Then $P_A + P_B = \frac{1}{2} + \frac{1}{2} = 1$. This property is called *normalization*.

c / The sum of the two dice is a random variable with possible values running from 2 to 12.



d / The histogram shows the probabilities of the various outcomes when rolling two dice.



e / A probability distribution for height of human adults. (Not real data.)

10.1.2 Continuous random variables

When numerical values are assigned to outcomes, the result is called a *random variable*. The sum of the rolls of two dice is a random variable, and we can assign probabilities to the different results. For example, the probability of rolling 2 is 1/36, since the probability of getting a 1 on the first die is 1/6, and similarly for the second die. All of the relevant information about probabilities can be summarized by the discrete function shown in figure d.

But when a random variable is continuous rather than discrete, we usually cannot make a useful graph of the probabilities, because the probability of any particular real number is typically zero. For example, there is zero probability that a person's height h will be 160 cm, since there are infinitely many possible results that are close to that value, such as 159.999999999999996876876587658465436 cm. What *is* useful to talk about is the probability that h will be *less than* a certain value. The probability of $h < 160$ cm is about 0.5. In general, we define the *cumulative probability distribution* $P(x)$ of a random variable to be the probability that the variable is less than or equal to x. We can then define the *probability distribution* of the variable to be

$$D(x) = P'(x). \tag{1}$$

Figure e shows an approximate probability distribution for human height. Suppose we want to know the probability that our random variable lies within the range from a to b. This is $P(b) - P(a)$. By the fundamental theorem of calculus, this can be calculated from the definite integral of the distribution,

$$P(b) - P(a) = \int_a^b D(x)\ \mathrm{d}x. \tag{2}$$

That is, areas under the probability distribution correspond to probabilities. If the random variable has some units, say centimeters, then the units of the probability distribution D are the inverse of those units, e.g., cm^{-1} in our example. In this example, D can be interpreted as the probability *per centimeter*. A *uniform* distribution is one for which D is a constant throughout the range of possible values of x.

An extremely common bell-shaped probability distribution is

$$D(x) = \frac{2}{\sqrt{\pi}} e^{-x^2},$$

called the "normal" or "Gaussian" distribution, which we encountered in section 8.7.1, p. 195.

If there are definite lower and upper limits L and U for the possible values of the random variable, then normalization requires that

$$1 = \int_L^U D(x)\ \mathrm{d}x. \tag{3}$$

The average $\bar{x}$ of a variable that takes on one of two discrete values with equal probability is $(x_1 + x_2)/2$, which is the same as $x_1 P_1 + x_2 P_2$. Generalizing this to a continuous random variable, we have

$$\bar{x} = \int_L^U x D(x) \, \mathrm{d}x. \tag{4}$$

The average is also known as the *mean*, *expectation*, or the *expected value* of x.

The standard deviation σ_x of a random variable x is a measure of how much it varies around its average value. The symbol σ is the lowercase Greek "sigma." (Recall that uppercase sigma is Σ.) The standard deviation of a continuous random variable is defined by

$$\sigma_x = \sqrt{\int_L^U (x - \bar{x})^2 D(x) \, \mathrm{d}x}. \tag{5}$$

10.1.3 One variable related to another

It often happens that one random variable y is defined by some function of some other random variable x. In an experiment, for example, one may measure x directly, and the value of x is a random variable because of the finite precision of the measurement. If one calculates the result of the experiment using some function $y(x)$, then the result is also a random variable. Let the corresponding probability distributions and cumulative probability distributions be D_x, D_y, and let P be the cumulative probability for a given x or y. Then D_y can be determined from D_x by the chain rule:

$$\begin{aligned}
D_y &= \frac{\mathrm{d}P}{\mathrm{d}y} && \text{[definition of } D\text{]} \\
&= \frac{\mathrm{d}P}{\mathrm{d}x} \cdot \frac{\mathrm{d}x}{\mathrm{d}y} && \text{[chain rule]} \\
&= D_x \cdot \frac{\mathrm{d}x}{\mathrm{d}y} && \text{[definition of } D\text{]} \\
&= \frac{D_x}{y'(x)} && \text{[derivative of the inverse of a function]}
\end{aligned}$$

A random goblin — Example 1

Often in computer simulations or games one wants to produce a random number with some desired distribution. For example, in a fantasy adventure game, we might wish to generate an opponent such as a goblin whose strength statistic y is distributed according to some bell-shaped curve D_y with a given mean and standard deviation. The random number generators supplied in computer programming libraries usually output a number x with a uniform distribution from 0 to 1, so that $D_x = 1$. We then have $y'(x) = 1/D_y$. Integrating both sides of this equation allows us to find a function $y(x)$ that determines the strength of the goblin.

10.2 Economics

▷Box 10.1 **Applications to economics**

The following is an index of applications of calculus to economics that occur throughout this book.

p.	*application*	
18	marginal rate of substitution	derivative
59	economic order quantity	extrema
106	the Laffer curve	Rolle's theorem
115	supply and demand	intermediate value theorem

In 1882, at the age of 46, William Stanley Jevons went swimming in the ocean and drowned. As a pioneer of classical economics, Jevons developed mathematical models that treated humans as rational actors seeking to maximize their happiness. His choice to go swimming that day was presumably based on the fact that swimming would cause him to be happy, and on the conscious or unconscious expectation that his risk of death would be low. But how do we define "rational" and "happiness" mathematically? Believe it or not, economists did produce definitions of these ideas, but in the process the word "happiness" changed to "utility," and the concepts morphed into forms that were very different from their original meanings. They are central to modern economics.

A 1947 paper by John von Neumann and Oskar Morgenstern (VNM) introduces four axioms defining rationality, which I'll describe here in English rather than equations:

1. Preferences are consistent.
2. Preferences are transitive: if you like outcome A more than B, and B more than C, then you like A more than C.
3. No outcome is infinitely good or bad. For example, if Jevons had believed that death was infinitely bad, he might have been unwilling to accept *any* risk of drowning. (Cf. example 11, p. 113.)
4. A preference for A over B holds regardless of whether some other outcome exists. For example, if you like Bach more than bebop, this is true regardless of whether it rains.

VNM prove that if these axioms hold, it is possible to assign a real number $u(x)$, called the utility function, to any outcome x such that a rational actor always maximizes the expected value of u as defined by equation (4), p. 217. The utility function can be rescaled or have a constant added to it, but is otherwise unique.

Although I've described this in terms of human preferences, the axioms may fail for humans or hold for non-humans. It only matters if the actor behaves *as if* it were acting rationally, as defined by the axioms. Milton Friedman writes:

> I suggest the hypothesis that the leaves [on a tree] are positioned as if each leaf deliberately sought to maximize the amount of sunlight it receives, given the position of its neighbors, as if it knew the physical laws determining the amount of sunlight that would be received in various positions and could move rapidly or instantaneously

from any one position to any other desired and unoccupied position.

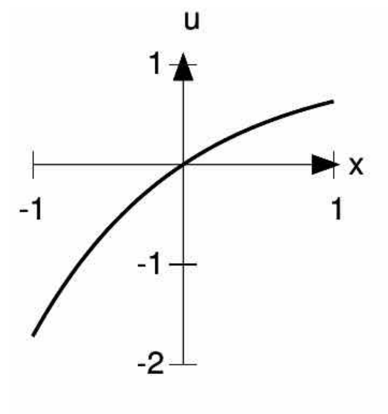
Daniel Kahneman, on the other hand, won the Nobel prize for his work showing that humans often violate the VNM definition of rationality, but in ways that can be described scientifically. For instance, he showed in experiments that subjects were willing to pay one price for a trinket such as a mug, but that if they were given the mug, they demanded a different and systematically higher price to sell it. This violates axiom 1. Axiom 1 was implicitly assumed in the description of the indifference curve in example 2, p. 18.

Playing the lottery Example 2

Joe is broke and homeless. He currently has an amount of money $x = 0$. Joe's utility function is given by

$$1 - e^{-x},$$

where x is in some appropriate units such as thousands of dollars. The shape of this function is shown in figure f. It is concave down, which is a feature that is almost always realistic for a utility function that depends on how much money someone has. If Joe is broke and gains \$10, he's really happy, whereas if Bill Gates saw a \$10 bill on the sidewalk, he probably wouldn't bother to bend over and pick it up.



f / The utility function of example 2.

Joe knows of a lottery in which each player receives a random amount of money uniformly distributed on the interval from 0 to 1. What price L should Joe be willing to pay for the lottery ticket, if he has the opportunity to borrow the price from his mother?

If Joe enters and receives the minimum payout of 0, he will have $x = -L$, i.e., he will be in debt to his mother for the price of the ticket and have nothing to show for it. If he gets the maximum reward of 1, he will have $x = 1 - L$. Since this interval has width 1, and the result is uniformly distributed, normalization requires that $D(x) = 1$ within the interval. We find his expected utility.

$$\begin{aligned}\bar{u} &= \int_{-L}^{1-L} u(x)D(x)\,\mathrm{d}x = \int_{-L}^{1-L} (1 - e^{-x})\,\mathrm{d}x \\ &= x + e^{-x}\Big]_{-L}^{1-L} = 1 - \left(1 - e^{-1}\right) e^{L}\end{aligned}$$

Joe's current utility function is $u(0) = 0$, so it is rational for him to pay any amount L that gives him $\bar{u} > 0$. The result is

$$L < -\ln\left(1 - e^{-1}\right) \approx 0.46.$$

If Joe's utility function had been $u(x) = x$, then he should have been willing to pay 0.5 units of money for a chance to win between 0 and 1 units. But because his utility function is nonlinear, he is willing to pay less than that; he is risk-averse.

10.3 Physics

▷Box 10.2 **Applications to physics**

The following is an index of applications of calculus to physics that occur throughout this book.

p.	*application*	
22	velocity	derivative
75	nuclear stability	extrema
83	acceleration	2nd derivative
88	Newton's 2nd law	2nd derivative
89	jerk and damage	3rd derivative
158	lever	related rates
161	pulley	implicit differentiation
187	work	definite integral
188	constant-acceleration motion	indefinite integral

A *conservation law* is a physical law stating that the total amount of a certain quantity stays constant. (This usage of "conservation" doesn't have the usual connotation of not using something up. In this context, the word implies that you *couldn't* use it up if you tried, because the total amount can't go down!) Some important examples of conserved quantities are mass, energy,[1] momentum, electric charge, and angular momentum (a measure of rotational motion). Conservation laws play a central role in physics. They are more fundamental than Newton's laws of motion. For example, a ray of light can be described by conservation of energy, but we get nonsense if we try to apply Newton's laws to it ($m = 0$, so we can't compute $a = F/m$).

Calculus deals with rates of change and the accumulation of change, so it would seem to have no application to variables that are guaranteed never to change! But conserved quantities can be transferred or transformed at some rate. For example, we estimated in example 9, p. 53, that hiking burns about 200 calories per hour. The calorie is a unit of energy.[2] This number represents the rate at which food energy is being transformed into other forms of energy such as body heat. For each conserved quantity, it's of interest to define a name, symbol, and unit for its rate of transfer or transformation. We then have two variables, which are related to one another as integral and derivative with respect to time. In the following table, the conserved quantity is given on top along with its symbol and SI unit. Its derivative is the variable below.

mass	energy	momentum	angular momentum	electric charge
m	E	p	L	q
kg	joule, J	N·s	N·m·s	coulomb, C
	power	force	torque	current
	P	F	τ	I
kg/s	watt, W	newton, N	N·m	ampere, A

Since the SI unit of time is the second (s), we have the following implied relationships between some of the units: W=J/s and A=C/s.

The *definitions* of the conserved quantities are ultimately operational definitions, meaning definitions that state the operations needed in order to *measure* them. This may seem unsatisfactory, but history has shown that every attempt at a "pure" conceptual or mathematical definition has had to be revised. We can however

[1] According to Einstein's famous $E = mc^2$, mass and energy are equivalent or interconvertible, so they aren't separately conserved. Their separate conservation is however a good approximation in ordinary life, where relativistic effects are negligible.

[2] Food calories are actually *kilo*calories, 1 kcal=1000 cal. The SI unit is not the calorie but the joule.

give rough conceptual definitions that are valid within the field of mechanics, i.e., the study of material objects:

Mass is a measure of inertia. How hard is it to change the motion of a certain object?

Momentum is a measure of the motion of an object. Suppose our object hits another object, the "target." Knowing the momentum allows us to predict how strongly a standard target will recoil. Momentum has a direction in space.

Energy comes in various forms such as kinetic energy (energy of motion), heat (which is random motion at the atomic level), and electrical energy (such as the chemical energy in food). Energy has no direction.

Box 10.3 gives some examples of equations for conserved quantities.

Energy of an accelerating car Example 3

▷ A car of mass m starts moving from rest with a constant acceleration a. If the speed is small enough, then air resistance is negligible, and the power required from the engine at time t is

$$P = kma^2t,$$

where the unitless fudge factor k accounts for inefficiency of the engine and frictional heating in the tires, and is assumed to be constant. Find the energy expended by burning gas as a function of time.

▷ Because the power isn't constant, we can't simply multiply "the" power by the time t. The integral is needed here as the correct generalization of multiplication (section 8.4.1, p. 186).

$$\begin{aligned} E &= \int P\,\mathrm{d}t && \text{[integral-derivative relationship of } E \text{ and } P\text{]} \\ &= \int kma^2t\,\mathrm{d}t \\ &= kma^2\int t\,\mathrm{d}t \\ &= \frac{1}{2}kma^2t^2 && \text{[let initial energy consumption=0]} \end{aligned}$$

For motion with constant acceleration, $v = at + v_o$, where $v_o = 0$ here because the car starts from rest. The result can therefore be rewritten as $(1/2)kmv^2$. The factor $(1/2)mv^2$ is called the kinetic energy of the car. If the car was perfectly efficient, we would have $k = 1$, and all the energy expended would go into kinetic energy, rather than frictional heating.

▷Box 10.3 **Examples of equations for conserved quantities**

Let a material object of mass m be moving at a velocity v that is small compared to the speed of light. Then experiments show that its momentum and kinetic energy are approximately $p = mv$ and $E = (1/2)mv^2$.

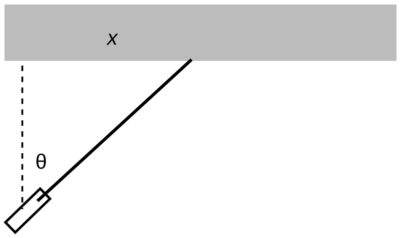
If a ray of light has energy E, then its momentum is $p = E/c$, where c is the speed of light. This momentum is too small to matter in everyday life.

If a material object moves at a speed that is *not* small compared to c, then it has $p = mv/\sqrt{1 - v^2/c^2}$.

Let a ring with mass m and radius r rotate about its own axis so that each point on it moves at speed v. Then its angular momentum is $\pm mvr$, with the sign indicating the direction of rotation.

Problems

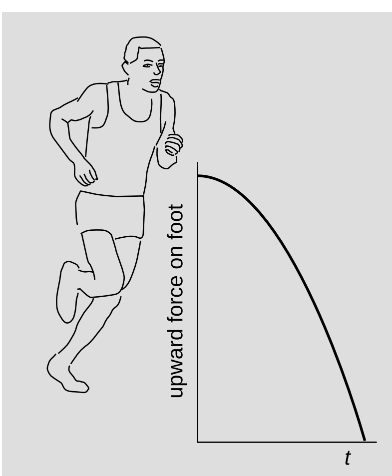
a1 A computer language will typically have a built-in subroutine that produces a fairly random number that is equally likely to take on any value in the range from 0 to 1. Find the standard deviation. $\checkmark$



Problem a2.

a2 A laser is placed one meter away from a wall, and spun on the ground to give it a random direction, but if the angle θ shown in the figure doesn't come out in the range from 0 to $\pi/2$, the laser is spun again until an angle in the desired range is obtained.
(a) Find the probability distribution D_θ of the variable θ.
(b) Using the technique described in section 10.1.3 on p. 217, find the probability distribution D_x of the distance x shown in the figure. $\checkmark$

a3 A computer language will typically have a built-in subroutine that produces a fairly random number that is equally likely to take on any value in the range from 0 to 1. If you take the absolute value of the difference between two such numbers, the probability distribution is of the form $D(x) = k(1 - x)$. (a) Find the value of the constant k that is required by normalization. $\checkmark$
(b) Find the average value of x. $\checkmark$
(c) Find the standard deviation. $\checkmark$



Problem c1.

c1 Scientists in Daniel Lieberman's Skeletal Biology Lab at Harvard specialize in measuring the forces that act on a runner's body, which may help to improve coaching, reduce injuries, or provide scientific evidence about whether barefoot running is healthier than using running shoes. The graph in the figure shows a typical result for the *vertical* force as a function of time that acts between the runner's foot and a treadmill, for one portion of a stride cycle.

The initial time $t = 0$ is the one when the vertical force is at its greatest, shown in the drawing. At this time, the runner's body is about as low as it will get, and the vertical momentum is approximately zero.

The end of the graph, where the force goes to zero, is the time at which the runner's back toe leaves the ground and he becomes airborne for a fraction of a second.

The graph looks like a parabola, so let's model it as one, $F = b(1-t^2/\tau^2)-w$, where τ is the time at which the graph ends, and the $-w$ term accounts for gravity. (a) Infer the units of the constants b, τ, and w. (b) Find the runner's vertical momentum at $t=\tau$, i.e., the momentum with which he takes off into the air. (c) Check that your answer to part b has units that make sense. $\checkmark$

e1 In example 2, p. 219, we found the maximum amount that a person should be willing to pay for a lottery ticket, given a certain utility function. We assumed the utility function to be concave down, which is usually realistic, for the reasons discussed in the example. But there can also be cases where the utility function is concave up. Suppose that Sally has cancer and no health insurance. She can only survive if she gets expensive treatment, which she can't presently afford. A small amount of money does her very little good, except that it slightly reduces the amount she still needs to get together for the treatment. In this situation, it might make sense to posit a concave-up utility function, such as $u(x)=e^x-1$, in the notation of the previous example. Redo the example with this utility function. $\checkmark$

Answers and solutions

Solutions to homework problems

Solutions for chapter 1

Page 36, problem a5:
The graph fails the vertical line test: a vertical line can pass through more than one point on the graph, meaning that there can be more than one pressure for a given temperature. Therefore p is not a function of T.

If we were to interchange the axes of the graph, it would pass the vertical line test. Therefore T can be described as a function of p. For a given pressure, there is only one temperature.

Page 36, problem a6:
A line will not be a function when it fails the vertical line test, i.e., when the line itself is a vertical line. Such a line is a set of points for which x is a constant. The equation $(\ldots)x + (\ldots)y + (\ldots) = 0$ can only be reduced to $x =$ constant if the coefficient of y is zero.

Page 36, problem a7:
All of them pass the vertical line test except for $x = y^2$, which has two y values for every positive x value. E.g., for $x = 4$, a vertical line passes through both $y = 2$ and $y = -2$.

Page 36, problem a8:
We have a set of points that are included in the set, which are those for which the given polynomial is negative. The set of points that are not included are those for which the polynomial is zero or positive. There is an edge or boundary between these two sets, consisting of any points at which the polynomial is zero, i.e., the roots of the polynomial. We could use the quadratic formula to find these roots. But since $u = 0$ is clearly a root, it's simpler just to factor the polynomial into $u(u - 2)$, which tells us that the other root is 2. Clearly the set S must be either the interval $(0, 2)$ or everything that lies *outside* this interval. Checking $u = 1$, we see that it's the former possibility that holds. Thus a simpler description is $\mathrm{S} = \{u | u > 0 \text{ and } u < 2\}$.

Page 37, problem c1:
The derivative is a rate of change, so the derivatives of the constants 1 and 7, which don't change, are clearly zero. The derivative can be interpreted geometrically as the slope of the tangent line, and since the functions t and $7t$ *are* lines, their derivatives are simply their slopes, 1, and 7. All of these could also have been found using the formula that says the derivative of t^k is kt^{k-1}, but it wasn't really

necessary to get that fancy. To find the derivative of t^2, we can use the formula, which gives $2t$. One of the properties of the derivative is that multiplying a function by a constant multiplies its derivative by the same constant, so the derivative of $7t^2$ must be $(7)(2t) = 14t$. By similar reasoning, the derivatives of t^3 and $7t^3$ are $3t^2$ and $21t^2$, respectively.

Page 37, problem c2:
They are the same function. A function is a graph that satisfies the vertical-line property. Both functions have all the same points in their graphs, so the two definitions have defined the same graph, which is the same function.

Page 37, problem c3:
Let m be the national budget surplus. For a brief period in an economic boom during the Clinton administration, the U.S. federal government had a budget surplus, so m was positive. Later, the economy cooled down and m became negative again — which is its normal state in the modern era. At some point in time t, m had to change from being positive to being negative, so $m(t) = 0$. At that moment, m was decreasing, so $m(t) < 0$.

Page 37, problem d1:
The addition property of the derivative tells us that we can break this down into the sum of the derivatives $(3x^4)'$, $(-2x^2)'$, $(x)'$, and $(1)'$. The derivative of the final, constant term is zero by the constant property. Using the power rule and adding, we have $12x^3 - 4x + 1$.

Page 38, problem e1:
One of the properties of the derivative is that the derivative of a sum is the sum of the derivatives, so we can get this by adding up the derivatives of $3z^7$, $-4z^2$, and 6. The derivatives of the three terms are $21z^6$, $-8z$, and 0, so the derivative of the whole thing is $21z^6 - 8z$.

For the numerical check, let's use $z = 1$ and $\Delta z = 0.001$. Call the function f.

$$\frac{\mathrm{d}f}{\mathrm{d}z} = 13$$
$$\frac{\Delta f}{\Delta z} = \frac{5.0131 - 5.0000}{0.001} = 13.1$$

These agree well enough that it's unlikely that we've made an error such as a wrong sign or getting the wrong integer for one of the coefficients.

Page 38, problem e6:
The first thing that comes to mind is the function f defined by $f(x) = 7x$. Its graph would be a line with a slope of 7, passing through the origin. Any other line with a slope of 7 would work too, e.g., $7x + 1$ and $7x - 42$.

Page 40, problem i1:
This is exactly like problem e1, except that instead of explicit numerical constants like 3 and -4, this problem involves symbolic constants a, b, and c. The result is $2at + b$.

Page 42, problem m1:
When the vertical stretch factor r is a natural number, that means that the function rf can be written as $f + f + \ldots + f$, where the number of terms in the sum is r. By the addition property of the derivative, the derivative of rf is then $f' + f' + \ldots + f'$, which is the same as rf'. This is the vertical stretch property.

Page 43, problem n1:
If the width and length of the rectangle are t and u, and Rick is going to use up all his fencing material, then the perimeter of the rectangle, $2t + 2u$, equals L, so for a given width, t, the length is $u = L/2 - t$. The area is $a = tu = t(L/2 - t)$. The function only means anything realistic for $0 \le t \le L/2$, since for values of t outside this region either the width or the height of the rectangle would be negative. The function $a(t)$ could therefore have a maximum either at a place where $\mathrm{d}a/\,\mathrm{d}t = 0$, or at the endpoints of the function's domain. We can eliminate the latter possibility, because the area is zero at the endpoints.

To evaluate the derivative, we first need to reexpress a as a polynomial:

$$a = -t^2 + \frac{L}{2}t.$$

The derivative is

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -2t + \frac{L}{2}.$$

Setting this equal to zero, we find $t = L/4$, as claimed.

Page 43, problem n2:
Since polynomials don't have kinks or endpoints in their graphs, the maxima and minima must be points where the derivative is zero. Differentiation bumps down all the powers of a polynomial by one, so the derivative of a third-order polynomial is a second-order polynomial. A second-order polynomial can have at most two real roots (values of t for which it equals zero), which are given by the quadratic formula. (If the number inside the square root in the quadratic formula is zero or negative, there could be less than two real roots.) That means a third-order polynomial can have at most two maxima or minima.

Page 44, problem r1:
The approximation we're going to use is

$$\frac{\mathrm{d}y}{\mathrm{d}x} \approx \frac{\Delta y}{\Delta x}.$$

Since we want an answer valid to three decimal places, it might be reasonable to try a Δx value such as 0.0001, since that's a lot smaller than 10^{-3}. We then have:

$$\frac{\Delta y}{\Delta x} = \frac{1/(1-0.0001) - 1/(1-0)}{0.0001 - 0} = 1.00010$$

It looks like we're getting 1 as our answer. To see if the result is really valid to three decimal places, we can try making Δx smaller, and see how much the result changes. With $\Delta x = 10^{-5}$, we get 1.00001. The change is in the fifth decimal place, so it looks like the first three decimal places are correct.

Page 45, problem s1:
(a) We have

$$\begin{aligned}
\Delta y &\approx \frac{\mathrm{d}y}{\mathrm{d}x}\Delta x \\
&= nkx^{n-1}\Delta x \\
\frac{\Delta y}{y} &\approx n\frac{\Delta x}{x}
\end{aligned}$$

(b) Here $n = 2$, so a relative error of 0.1% in the length will cause a 0.2% error in the area.

Page 45, problem s2:
Thinking of the rocket's height as a function of time, we can see that goal is to measure the function at its maximum. The derivative is zero at the maximum, so the error incurred due to timing is approximately zero. She should not worry about the timing error too much. Other factors are likely to be more important, e.g., the rocket may not rise exactly vertically above the launchpad.

Solutions for chapter 2

Page 69, problem e1:
Reexpressing $\sqrt[3]{x}$ as $x^{1/3}$, the derivative is $(1/3)x^{-2/3}$.

Page 69, problem e2:
(a) Using the chain rule, the derivative of $(x^2+1)^{1/2}$ is $(1/2)(x^2+1)^{-1/2}(2x) = x(x^2+1)^{-1/2}$.
(b) This is the same as a, except that the 1 is replaced with an a^2, so the answer is $x(x^2+a^2)^{-1/2}$. The idea would be that a has the same units as x.
(c) This can be rewritten as $(a+x)^{-1/2}$, giving a derivative of $(-1/2)(a+x)^{-3/2}$.
(d) This is similar to c, but we pick up a factor of $-2x$ from the chain rule, making the result $ax(a-x^2)^{-3/2}$.

Page 70, problem e4:
The vertical stretch rule says that stretching a function $y(x)$ vertically to form a new function $ry(x)$ multiplies its derivative by r at the corresponding points. That is, if r is a constant, then $(ry)' = ry'$. To prove this using the product rule, we have

$$(ry)' = r'y + y'r.$$

But if r' is a constant, then $r' = 0$, so the first term is zero, and we have the claimed result.

Page 71, problem i2:
Let P be the point $(1, 1)$, and let Q lie on the graph at $x = 1 + \mathrm{d}x$. The slope of the line through P and Q is

$$\begin{aligned} \text{slope of line PQ} &= \frac{\Delta y}{\Delta x} \\ &= \frac{(1+\mathrm{d}x)^3 - 1}{(1+\mathrm{d}x) - 1} \\ &= \frac{3\,\mathrm{d}x + 3\,\mathrm{d}x^2 + \mathrm{d}x^3}{\mathrm{d}x} \end{aligned}$$

Discarding the $\mathrm{d}x^2$ and $\mathrm{d}x^3$ terms, this becomes 3, which is the same as the result we got by doing limits.

Page 71, problem i3:
This would be a horrible problem if we had to expand this as a polynomial with 101 terms, as in chapter 1! But now we know the chain rule, so it's easy. The derivative is

$$\left[100(2x+3)^{99}\right][2],$$

where the first factor in brackets is the derivative of the function on the outside, and the second one is the derivative of the "inside stuff." Simplifying a little, the answer is $200(2x + 3)^{99}$.

Page 71, problem i4:
Applying the product rule, we get

$$100(x+1)^{99}(x+2)^{200} + 200(x+1)^{100}(x+2)^{199}.$$

(The chain rule was also required, but in a trivial way — for both of the factors, the derivative of the "inside stuff" was one.)

Page 71, problem i5:
The chain rule gives

$$\frac{\mathrm{d}}{\mathrm{d}x}((x^2)^2)^2 = 2((x^2)^2)(2(x^2))(2x) = 8x^7,$$

which is the same as the result we would have gotten by differentiating x^8.

Page 71, problem i6:
Converting these into Leibniz notation, we find

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}g}{\mathrm{d}h}$$

and

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}g}{\mathrm{d}h} \cdot h.$$

To prove something is not true in general, it suffices to find one counterexample. Suppose that g and h are both unitless, and x has units of seconds. The value of f is defined by the output of g, so f must also be unitless. Since f is unitless, $\mathrm{d}f/\mathrm{d}x$ has units of inverse seconds ("per second"). But this doesn't match the units of either of the proposed expressions, because they're both unitless. The correct chain rule, however, works. In the equation

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}g}{\mathrm{d}h} \cdot \frac{\mathrm{d}h}{\mathrm{d}x},$$

the right-hand side consists of a unitless factor multiplied by a factor with units of inverse seconds, so its units are inverse seconds, matching the left-hand side.

Page 74, problem p1:
We can make life a lot easier by observing that the function $s(f)$ will be maximized when the expression inside the square root is minimized. Also, since f is squared every time it occurs, we can change to a variable $x = f^2$, and then once the optimal value of x is found we can take its square root in order to find the optimal f. The function to be optimized is then

$$a(x - f_o^2)^2 + bx.$$

Differentiating this and setting the derivative equal to zero, we find

$$2a(x - f_o^2) + b = 0,$$

which results in $x = f_o^2 - b/2a$, or

$$f = \sqrt{f_o^2 - b/2a},$$

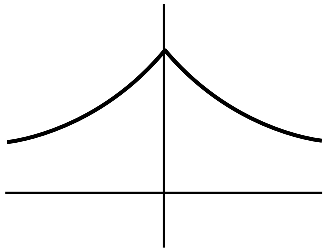
(choosing the positive root, since f represents a frequencies, and frequencies are positive by definition). Note that the quantity inside the square root involves the square of a frequency, but then we take its square root, so the units of the result turn out to be frequency, which makes sense. We can see that if b is small, the second term is small, and the maximum occurs very nearly at f_o.

There is one subtle issue that was glossed over above, which is that the graph on page 74 shows *two* extrema: a minimum at $f = 0$ and a maximum at $f > 0$. What happened to the $f = 0$ minimum? The issue is that I was a little sloppy with the change of variables. Let I stand for the quantity inside the square root in the original expression for s. Then by the chain rule,

$$\frac{\mathrm{d}s}{\mathrm{d}f} = \frac{\mathrm{d}s}{\mathrm{d}I} \cdot \frac{\mathrm{d}I}{\mathrm{d}x} \cdot \frac{\mathrm{d}x}{\mathrm{d}f}.$$

We looked for the place where $\mathrm{d}I/\mathrm{d}x$ was zero, but $\mathrm{d}s/\mathrm{d}f$ could also be zero if one of the other factors was zero. This is what happens at $f = 0$, where $\mathrm{d}x/\mathrm{d}f = 0$.

Page 78, problem t1:
The graph looks like this:



Clearly it has a kink in it. No matter how far we zoom in, the kink will never look like a line. The function is not differentiable at $x = 0$.

Page 78, problem t2:
The function $f(x) = 1/\sin x$ can be written as a composition $f(x) = g(h(x))$ of the functions $g(x) = 1/x$ and $h(x) = \sin x$. We don't have to recall anything about the sine function, h, except that it looks like a sine wave, so that it's clearly continuous and differentiable everywhere. The function g, on the other hand, is discontinuous at 0, so it will be discontinuous at any x such that $\sin x = 0$, and f will also be discontinuous in these places. The relevant values of x are $\{\ldots, -2\pi, -\pi, 0, \pi, 2\pi, \ldots\}$. Since f is discontinuous at these points, it is also nondifferentiable there, because discontinuity implies nondifferentiability.

Page 78, problem t3:
A cusp will occur if both branches are vertical at $x = 0$, i.e., if f' blows up there.

For positive values of x, the definition of f is the same as x^p, so by the power rule $f' = px^{p-1}$. For negative x, the horizontal flip property of the derivative (p. 16) tells us that f' equals minus the value of the derivative at the corresponding point on the right.

For $p < 1$, the derivative blows up, and f has a cusp.

If f is to be differentiable at $x = 0$, then it can't have a kink. By the symmetry property described above, this requires that $f'(0) = 0$. This occurs if $p > 1$. The function is nondifferentiable when $p \leq 1$.

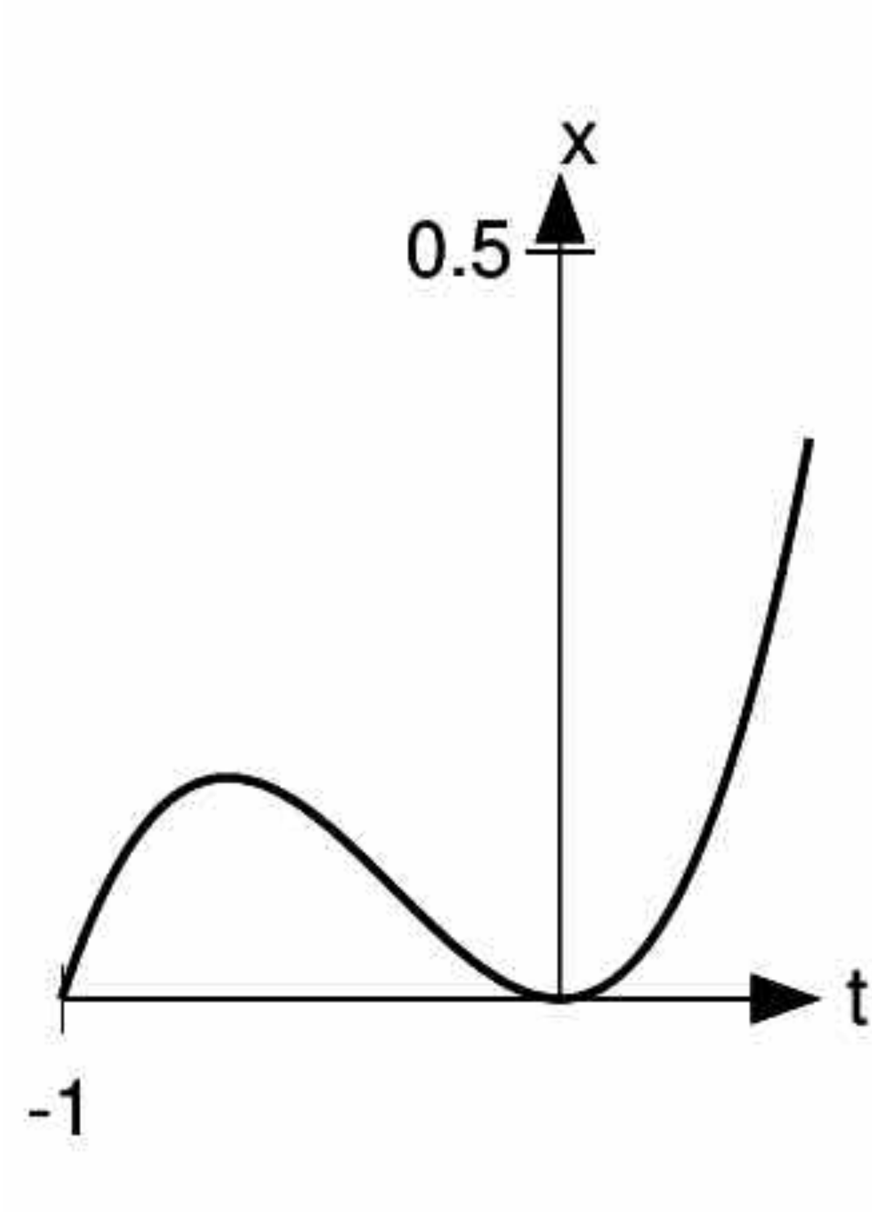
Page 80, problem y1:
We can derive a three-factor product rule by grouping the three factors into two factors, and then applying the two-factor rule.

$$\begin{aligned}(fgh)' &= [(fg)h]' \\ &= (fg)'h + h'fg \\ &= (f'g + g'f)h + h'fg \\ &= f'gh + g'hf + h'fg\end{aligned}$$

Solutions for chapter 3

Page 92, problem a1:
The first derivative is $12z^3 - 8z$. Differentiating a second time, we get $36z^2 - 8$.

Page 92, problem c1:
The first derivative is $3t^2 + 2t$, and the second is $6t + 2$. Setting this equal to zero and solving for t, we find $t = -1/3$. Looking at the graph, it does look like the concavity is down for $t < -1/3$, and up for $t > -1/3$.



Page 92, problem c2:
Since f, g, and s are smooth and defined everywhere, any extrema they possess occur at places where their derivatives are zero. The

converse is not necessarily true, however; a place where the derivative is zero could be a point of inflection. The derivative is additive, so if *both* f and g have zero derivatives at a certain point, s does as well. Therefore in most cases, if f and g both have an extremum at a point, so will s. However, it could happen that this is only a point of inflection for s, so in general, we can't conclude anything about the extrema of s simply from knowing where the extrema of f and g occur.

Going the other direction, we certainly can't infer anything about extrema of f and g from knowledge of s alone. For example, if $s(x) = x^2$, with a minimum at $x = 0$, that tells us very little about f and g. We could have, for example, $f(x) = (x-1)^2/2 - 2$ and $g(x) = (x+1)^2/2 + 1$, neither of which has an extremum at $x = 0$.

Solutions for chapter 4

Page 121, problem a1:

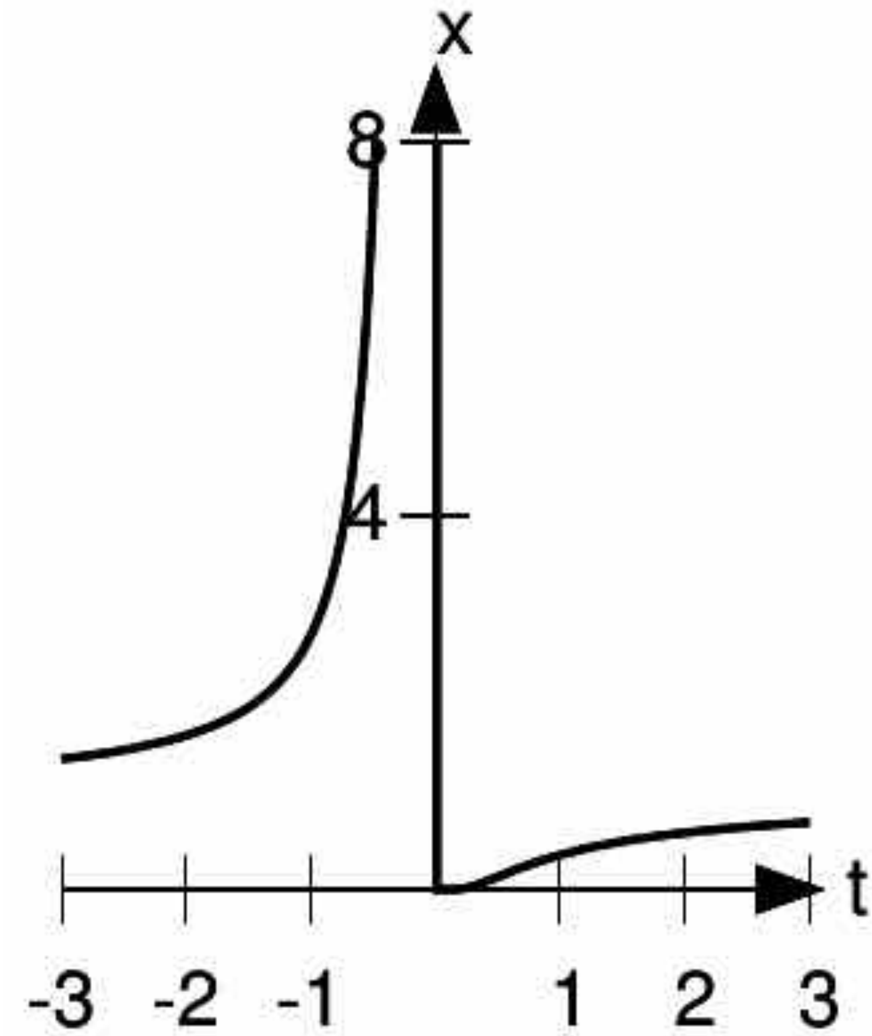
x	$\sqrt{x+1} - \sqrt{x-1}$
1000	.032
$1000,000$	0.0010
$1000,000,000$	0.00032

The result is getting smaller and smaller, so it seems reasonable to guess that the limit is zero.

Page 121, problem a2:
If R_1 is finite and R_2 approaches infinity, then $1/R_2$ is approaches zero. $1/R_1 + 1/R_2$ approaches $1/R_1$, and the combined resistance R approaches from R_1. Physically, the second pipe is blocked or too thin to carry any significant flow, so it's as though it weren't present.

If R_1 is finite and R_2 gets very small, then $1/R_2$ gets very big, $1/R_1 + 1/R_2$ is dominated by the second term, and the result is basically the same as R_2. It's so easy for water to flow through R_2 that R_1 might as well not be present. In the context of electrical circuits rather than water pipes, this is known as a short circuit.

Page 121, problem c1:
The shape of the graph can be found by considering four cases: large negative x, small negative x, small positive x, and large positive x. In these four cases, the function is respectively close to 1, large, small, and close to 1.



The four limits correspond to the four cases described above.

Page 123, problem c8:
For x approaching $\pm\infty$, the x^2 term dominates, and the function approaches zero. Therefore the function has a horizontal asymptote at zero.

Each root of the polynomial in the denominator will correspond to a vertical asymptote. These roots can be determined from the quadratic formula, which contains the square root of $b^2 - 4ac$, called the discriminant. If the discriminant is greater than zero, then there will be two asymptotes, corresponding to the positive and negative roots of the discriminant. If the discriminant is zero, then there will be only one real root and one vertical asymptote. If the discriminant is negative, then there are no real roots and no vertical asymptotes.

Page 123, problem c9:
It has a vertical asymptote where the denominator blows up, at $x = -1$. It has horizontal asymptotes at $y = 1$, since in the limits as x approach $\pm\infty$, the numerator and denominator are dominated by the x^7 terms, and the constant terms become unimportant.

Page 123, problem c10:
The function

$$f(x) = \left(\frac{x^2+1}{x^2+2} - \frac{x^2+3}{x^2+4}\right)^{-1}$$

is not given in the form of a rational function, and the most straightforward thing to do here would be simply to change it into that form. Before we do that, however, we could look for values of x at which the quantity inside the parentheses would go to zero; these would be the vertical asymptotes. Setting the denominator equal to zero gives $(x^2+1)(x^2+4) = (x^2+2)(x^2+3)$, which simplifies to $4 = 6$. There are no solutions, and therefore the function has no vertical asymptotes.

Going ahead and recasting it as a rational function, we first need to put the two terms over a common denominator. This gives

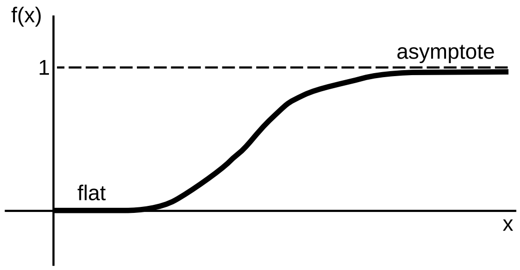
$$f(x) = \left(\frac{(x^2+1)(x^2+4) - (x^2+2)(x^2+3)}{(x^2+2)(x^2+4)}\right)^{-1},$$

which simplifies to

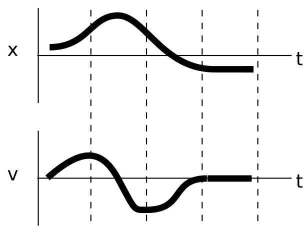
$$\begin{aligned} f(x) &= \left(\frac{-2}{(x^2+2)(x^2+4)}\right)^{-1} \\ &= -\frac{1}{2}(x^2+2)(x^2+4). \end{aligned}$$

We now see that the exotic-looking function was in fact just a polynomial in disguise. Polynomials don't have horizontal or vertical asymptotes.

Page 123, problem e1:
Clearly f will be a non-decreasing function and will asymptotically approach 1 as x approaches infinity. We can also say something about the value of $f'(0)$. Bounty hunting is a nasty, dirty, dangerous business that requires a significant up-front investment. Therefore we don't expect any bounty hunters to become active unless x is high enough to give them some expectation of making a profit, and we expect both $f(0) = 0$ and $f'(0) = 0$, and the function should be essentially zero until it starts to rise at some finite value of x.



Page 124, problem g1:



Page 124, problem k1:
If f'' is continuous and sometimes positive and sometimes negative, then by the intermediate value theorem there is a point where $f''(x) = 0$. (This is the part of the argument that fails for a function on the rationals.) Furthermore, we must have some such x at which f'' changes sign, and this is by definition a point of inflection.

Solutions for chapter 5

Page 138, problem a1:
A point on the unit circle has coordinates $(x, y) = (\cos\theta, \sin\theta)$, where θ is the angle measured counterclockwise from the x axis. If we want both sine and cosine to be negative, then we need a point on the unit circle that lies in the third quadrant, excluding the points that coincide with the axes. That means $\theta \in (\pi, 3\pi/2)$.

Page 139, problem c1:
By the chain rule, the result is $2/(2t + 1)$.

Page 139, problem c2:
We need to put together three different ideas here: (1) When a function to be differentiated is multiplied by a constant, the constant just comes along for the ride. (2) The derivative of the sine is the cosine. (3) We need to use the chain rule. The result is $ab\cos(bx+c)$.

Page 139, problem c3:
The derivative of e^{7x} is $e^{7x} \cdot 7$, where the first factor is the derivative of the outside stuff (the derivative of a base-e exponential is just the same thing), and the second factor is the derivative of the inside stuff. This would normally be written as $7e^{7x}$.

The derivative of the second function is $e^{e^x}e^x$, with the second exponential factor coming from the chain rule.

Page 139, problem c4:
To find a maximum, we take the derivative and set it equal to zero. The whole factor of $2v^2/g$ in front is just one big constant, so it comes along for the ride. To differentiate the factor of $\sin\theta\ \cos\theta$, we need to use the chain rule, plus the fact that the derivative of sin is cos, and the derivative of cos is $-\sin$.

$$\begin{aligned}0 = \frac{2v^2}{g}(\cos\theta\ \cos\theta + \sin\theta(-\sin\theta))\\ 0 = \cos^2\theta - \sin^2\theta\\ \cos\theta = \pm\sin\theta\end{aligned}$$

We're interested in angles between, 0 and 90 degrees, for which both the sine and the cosine are positive, so

$$\begin{aligned}\cos\theta = \sin\theta\\ \tan\theta = 1\\ \theta = 45\,°.\end{aligned}$$

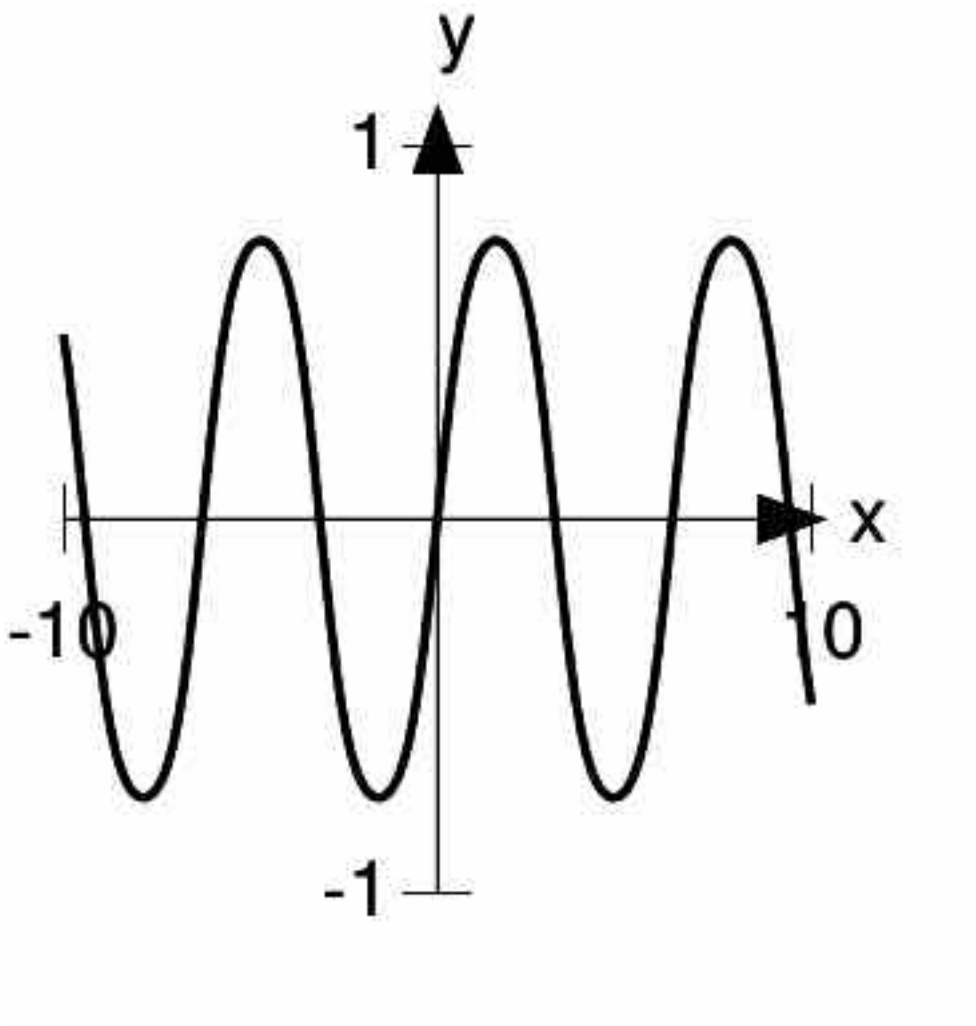
To check that this is really a maximum, not a minimum or an inflection point, we could resort to the second derivative test, but we know the graph of $R(\theta)$ is zero at $\theta = 0$ and $\theta = 90\,°$, and positive in between, so this must be a maximum.

Page 139, problem c5:
Since I've advocated not memorizing the quotient rule, I'll do this one from first principles, using the product rule.

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\theta}\tan\theta\\ &\quad= \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\sin\theta}{\cos\theta}\right)\\ &\quad= \frac{\mathrm{d}}{\mathrm{d}\theta}\left[\sin\theta\,(\cos\theta)^{-1}\right]\\ &\quad= \cos\theta\,(\cos\theta)^{-1} + (\sin\theta)(-1)(\cos\theta)^{-2}(-\sin\theta)\\ &\quad= 1 + \tan^2\theta\end{aligned}$$

(Using a trig identity, this can also be rewritten as $\sec^2\theta$.)

Page 139, problem c6:
There are no kinks, endpoints, etc., so extrema will occur only in places where the derivative is zero. Applying the chain rule, we find the derivative to be $\cos(\sin(\sin x))\cos(\sin x)\cos x$. This will be zero if any of the three factors is zero. We have $\cos u = 0$ only when $|u| \ge \pi/2$, and $\pi/2$ is greater than 1, so it's not possible for either of the first two factors to equal zero. The derivative will therefore equal zero if and only if $\cos x = 0$, which happens in the same places where the derivative of $\sin x$ is zero, at $x = \pi/2 + \pi n$, where n is an integer.



Page 139, problem c7:
Taking the derivative and setting it equal to zero, we have $\left(e^x - e^{-x}\right)/2 = 0$, so $e^x = e^{-x}$, which occurs only at $x = 0$. The second derivative is $\left(e^x + e^{-x}\right)/2$ (the same as the original function), which is positive for all x, so the function is everywhere concave up, and this is a minimum.

Page 141, problem f1:
Let us first pause to mourn the loss of this perfectly good bottle of beer, and to vow that such a thing must never be allowed to happen again.
(a) Since T has units of degrees, both terms on the right-hand side must also have units of degrees. The first term on the right is a, so a has units of degrees. The second term consists of b multiplied by an exponential. The exponential is unitless, so b must have units of degrees. The input to the exponential must be unitless as well, so c must have units of inverse seconds (s^{-1}).
(b) $\mathrm{d}T/\,\mathrm{d}t = bce^{-ct}$
On the left side, the units are what is implied by the original in-

terpretation of the Leibniz notation: we have a small change in temperature divided by a small change in time, so the units are degrees per second ($^\circ$/s). On the right, the units come from the factor bc, since the exponential is unitless. The units of bc are degrees multiplied by inverse seconds, ($^\circ$)(s^{-1}), and this matches what we had on the left-hand side. (c) In this limit, the the temperature approaches a, and the derivative approaches zero. It makes sense that the derivative goes to zero, since eventually the beer will be in thermal equilibrium with the air.
(d) Physically, a is the temperature of the air, b is the difference in temperature at $t=0$ between the air and the beer, and c measures how good the thermal contact is between the air and the beer — e.g., if the beer is in a styrofoam container, c will be small.

Solutions for chapter 6

Page 153, problem a1:
All five of these can be done using l'Hôpital's rule:

$$\begin{aligned}
&\lim_{s\to 1}\frac{s^3-1}{s-1}=\lim\frac{3s^2}{1}=3\\
&\lim_{\theta\to 0}\frac{1-\cos\theta}{\theta^2}=\lim\frac{\sin\theta}{2\theta}=\lim\frac{\cos\theta}{2}=\frac{1}{2}\\
&\lim_{x\to\infty}\frac{5x^2-2x}{x}=\lim\frac{10x-2}{1}=\infty\\
&\lim_{n\to\infty}\frac{n(n+1)}{(n+2)(n+3)}=\lim\frac{n^2+\ldots}{n^2+\ldots}=\lim\frac{2n+\ldots}{2n+\ldots}=\lim\frac{2}{2}=1\\
&\lim_{x\to\infty}\frac{ax^2+bx+c}{dx^2+ex+f}=\lim\frac{2ax+\ldots}{2dx+\ldots}=\lim\frac{2a}{2d}=\frac{a}{d}
\end{aligned}$$

In examples 2, 4, and 5, we differentiate more than once in order to get an expression that can be evaluated by substitution. In 4 and 5, $\ldots$ represents terms that we anticipate will go away after the second differentiation. Most people probably would not bother with l'Hôpital's rule for 3, 4, or 5, being content merely to observe the behavior of the highest-order term, which makes the limiting behavior obvious. Examples 3, 4, and 5 can also be done rigorously without l'Hôpit rule, by algebraic manipulation; we divide on the top and bottom by the highest power of the variable, giving an expression that is no longer an indeterminate form ∞/∞.

Page 153, problem a2:
Both numerator and denominator go to zero, so we can apply l'Hôpital's rule. Differentiating top and bottom gives $(\cos x - x\sin x)/(-\ln 2\cdot 2^x)$, which equals $-1/ln2$ at $x=0$. To check this numerically, we plug $x=10^{-3}$ into the original expression. The result is -1.44219, which is very close to $-1/ln2=-1.44269\ldots$.

Page 153, problem a3:
L'Hôpital's rule only works when both the numerator and the denominator go to zero.

Page 153, problem a4:
Applying l'Hôpital's rule once gives

$$\lim_{u\to 0} \frac{2u}{e^u - e^{-u}},$$

which is still an indeterminate form. Applying the rule a second time, we get

$$\lim_{u\to 0} \frac{2}{e^u + e^{-u}} = 1.$$

As a numerical check, plugging $u = 0.01$ into the original expression results in 0.9999917.

Page 153, problem a5:
L'Hôpital's rule gives $\cos t/1 \rightarrow -1$. Plugging in $t = 3.1$ gives -0.9997.

Solutions for chapter 7

Page 169, problem e1:
We have the same power law for differentials as for derivatives, so the result is $52B^{51}\,\mathrm{d}B$. Note that the answer is wrong without the $\mathrm{d}B$. If we think of differentials as "a little bit of...," then $\mathrm{d}(B^{52})$ means a tiny change in B^{52}. It can't equal $52B^{51}$, because $52B^{51}$ is not typically going to be tiny.

Page 169, problem e2:
As with derivatives, a constant factor just "comes along for the ride," so $\mathrm{d}(2000BC) = 2000\,\mathrm{d}(BC)$. We have the same product rule for differentials as for derivatives, so the result is $2000(B\,\mathrm{d}C + C\,\mathrm{d}B)$.

Page 169, problem e3:
We have the same chain rule for differentials as for derivatives. If k had been a function of some other variable t, and we'd been taking the derivative of $\sin k$ with respect to t, then we would have had $\cos k\,\mathrm{d}k/\,\mathrm{d}t$. For the differential we have simply $\cos k\,\mathrm{d}k$.

Page 169, problem e4:
Applying the sum rule and then the product rule, we have
$p\,\mathrm{d}b + b\,\mathrm{d}p + \mathrm{d}j$.

Page 170, problem g1:
Squaring both sides clears the square root.

$$y^2 = x^2 + 1$$

Implicit differentiation gives the following.

$$\begin{aligned} 2y\,\mathrm{d}y &= 2x\,\mathrm{d}x \\ \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{x}{y} \\ &= \frac{x}{\sqrt{x^2+1}} \end{aligned}$$

Page 170, problem i1:

$$\begin{aligned} &e^{x+y}\,\mathrm{d}x + xe^{x+y}(\mathrm{d}x+\mathrm{d}y) + \mathrm{d}y = 0 \\ &\left(e^{x+y} + xe^{x+y}\right)\mathrm{d}x + \left(xe^{x+y}+1\right)\mathrm{d}y = 0 \\ &\frac{\mathrm{d}y}{\mathrm{d}x} = -\left(\frac{1+x}{1+xe^{x+y}}\right)e^{x+y} \end{aligned}$$

Plugging in $x = 0$ and $y = 0$ gives $\mathrm{d}y/\,\mathrm{d}x = -1$.

Solutions for chapter 8

Page 197, problem a1:
The given equation

$$P_2 - P_1 = \rho g \Delta y$$

involves multiplication of a number ρ by a number $g\Delta y$. If ρ is not constant, then the proper way to generalize multiplication is through an integral.

$$P_2 - P_1 = \int_{y_1}^{y_2} \rho g\,\mathrm{d}y$$

Page 197, problem a2:
The two options proposed are:

$$\begin{aligned} \mathrm{PV} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-rT}f(t)\right) \\ \mathrm{PV} &= \int_0^T f(t)e^{-rt}\,\mathrm{d}t \end{aligned}$$

The units of the present value should be dollars.

The first proposed equation is nonsense based on units, because f has units of dollars/year, and its time derivative would therefore have units of dollars/year^2, not dollars.

The units of the second equation do make sense. The Leibniz notation for the integral is designed so that if you analyze the units and treat the integral sign as a sum, the units are what they look like they are. On the right-hand side, the units are (dollars/year) $\times$ years = dollars, which matches the units on the left-hand side. This doesn't prove that this equation is right, but it doesn't prove it wrong, either.

Page 197, problem a3:
The proposed relationships are:

$$I = k\frac{\mathrm{d}C}{\mathrm{d}t}$$
$$I = k\int_{t_1}^{t_2} C \ \mathrm{d}t$$

A derivative represents a rate of change, while an integral represents the accumulation of change. Based on these concepts, the first equation makes sense: the current tells us how fast our accumulated bill is adding up. The second one doesn't make sense conceptually.

Page 198, problem c2:
(a)

$$\int_0^1 x$$

This one is wrong because it's written ungrammatically. It's wrong without the dx, for the reasons explained on p. 180.

$$\int_0^1 x \ \mathrm{d}x$$

This one is correct.

$$\int_0^1 u \ \mathrm{d}u$$

This one is also correct. It doesn't matter that a different letter is used. The x or u is just a dummy variable.

(b) The correct way to notate this is $\int \left(x^2+1\right) \ \mathrm{d}x$, so that the differential dx is being multiplied by the whole expression. The notation $\int x^2 + 1 \ \mathrm{d}x$ makes it look like the dx is only multiplying the 1.

Page 198, problem e1:
We know that the derivative of e^x is e^x. Adding a constant doesn't matter, so two more possibilities are $e^x + 7$ and $e^x + 13$.

Page 198, problem e2:

$$\int x \ \mathrm{d}x = \frac{1}{2}x^2 + c$$

Differentiating the right-hand side gives $\frac{1}{2}(2x) = x$, which is correct. (The derivative of the constant term is zero.)

$$\int x^4 \ \mathrm{d}x = 4x^5 + c$$

Differentiating the right-hand side would give $20x^4$, which is wrong. The coefficient on the right should be 1/5, not 4.

$$\int e^x \, \mathrm{d}x = e^x + c$$

Differentiation gives e^x, which is right.

$$\int e^{2x} \, \mathrm{d}x = e^{2x} + c$$

Differentiation gives $2e^{2x}$, where the factor of 2 in front comes from the chain rule. The integral is wrong as written. It should have a factor of 1/2 in front.

$$\int x^{-1} \, \mathrm{d}x = x^0 + c$$

This is wrong. Raising something to the power 0 simply gives 1, so the right-hand side is $1 + c$, which is a constant. If we differentiate it, we get zero, not x^{-1}. As in example 7, p. 185, the correct integral is $\ln x + c$.

Page 200, problem i1:
First we put the integrand into the more familiar and convenient form cx^p, whose integral is $(c/(p+1))x^{p+1}$:

$$\sqrt{Bx\sqrt{x}} = B^{1/2}x^{3/4}$$

Applying the general rule, the result is $(4/7)B^{1/2}x^{7/4}$.

Page 201, problem n1:
(a) As described in the instructions above the problem, force has units of newtons (N). Since distance is measured in meters (m), the constant k must have units of N/m.
(b)

$$W = \int_0^b kx \, \mathrm{d}x = \frac{1}{2}kx^2\Big]_0^b = \frac{1}{2}kb^2$$

(c) As described in the instructions, work has units of N·m, so we need to check that the expression $(1/2)kb^2$ also has these units. The 1/2 is unitless. The constant k has units of N/m, and multiplying these units by meters squares does give N·m.

Solutions for chapter 9

Solutions for chapter 10

Photo credits

Cover Megyeri Bridge, (c) 2008, Tamás Mészöly, CC-BY-SA

?? *Rock climber:* Line art by B. Crowell, CC-BY-SA. Based on a photo by Jason McConnell-Leech, CC-BY-SA. **??** *Graph:* Redrawn by B. Crowell from a student research paper by Casey Johnson and Charlie Klonowski, Cal Poly Pomona. **25** *Vanishing point:* Wikimedia Commons user Zorba the Geek, CC-BY-SA. **26** *Simon Stevin:* Statue by Simonis, 19th century. Photo by Ad Meskens, attribution required. **27** *Leibniz:* Contemporary, public domain. **31** *Baseketball photo:* Wikimedia Commons user Reisio, public domain. **34** *Snake:* Wikipedia user DuSantos, CC-BY. **36** *Graph of Galileo data:* Seiff *et al.*, J. Geophys. Research 103 (1998) 22,857. **54** *Gears:* Jared C. Benedict, CC-BY-SA. **54** *Russian dolls:* Wikimedia Commons users Fanghong and Gnomz007, CC-BY-SA. **56** *Bucket brigade:* Agência Brasil, CC-BY. **58** *Dominoes:* Jochen Burghardt, CC-BY-SA. **61** *Cusp (caustic) in a teacup:* Wikipedia user Paul Venter, CC-BY-SA. **83** *Beer photo:* Wikipedia user Fgeerts, public domain. **88** *Isaac Newton:* Godfrey Kneller, 1702. **140** *Mercator projection:* Wikimedia Commons user Strebe, CC-BY-SA. **143** *Mortality graph:* Wikipedia user Uscitizenjason, CC-BY-SA. **155** *Cusp (caustic) in a teacup:* Wikipedia user Paul Venter, CC-BY-SA. **155** *Slide rule:* B. Crowell, CC-BY-SA. **159** *Photo of camshaft:* Wikipedia user Stahlkocher, CC-BY-SA. **169** *Talus cones:* Mark A. Wilson, public domain. **171** *Astroid:* Wikipedia user Joelholdsworth, CC-BY-SA. **171** *Sergel's Square fountain:* German Wikipedia user Stern, CC-BY-SA. **173** *Gauss:* C.A. Jensen (1792-1870). **174** *Riemann:* Contemporary, public domain. **192** *Tree rings:* Wikimedia commons user Arnoldius, CC-BY-SA. **195** *Error function:* Based on a drawing by J. Kemp and P. Strandmark, Wikipedia.. **201** *Jupiter:* Hubble Space Telescope, NASA, not copyrighted. **222** *Runner:* Wikimedia Commons user user Citizen59, CC-BY-SA.

Index